**Chairperson: Bob Wyatt, NW Natural**
**Treasurer: Fred Wolf, Legacy Site Services for Arkema**

November 13, 2009

Chip Humphrey
Eric Blischke
US Environmental Protection Agency, Region 10
805 SW Broadway, Suite 500
Portland, OR 97205

**Re: Draft Benthic Reanalysis Technical Memorandum for the Portland Harbor Superfund Site**

Chip and Eric:

The Portland Harbor draft baseline ecological risk assessment (BERA), hereafter referred to as the draft BERA, was compiled with oversight by the US Environmental Protection Agency (EPA) and submitted for review on September 2, 2009 (Windward 2009). Risks to benthic invertebrate communities from exposure to sediment-borne contaminants in the Lower Willamette River were assessed in the draft BERA. The approach to assess benthic community risks was developed and refined over a number of years based on EPA's formal comments on Lower Willamette Group (LWG) project deliverables, as well as workshops, conference calls, e-mails, and letters.

EPA provided additional advice and instructions concerning the benthic BERA between June 18 and August 28, 2009, which unfortunately came at a point in time when it was impossible to consider them for incorporation into the draft BERA without delaying its submittal. EPA directed the LWG to not modify the approach for evaluating benthic risk and to submit the draft BERA as soon as practicable (EPA 2009c). The LWG complied with that direction.

Once the draft BERA had been submitted, the LWG undertook a reanalysis of benthic risk following the advice and instructions provided by EPA in June, July, and August 2009. Those instructions changed benthic toxicity "hit" thresholds and the hit classifications assigned to the bioassay stations based on these hit thresholds, which resulted in minor changes to site-specific sediment quality guidelines. That reanalysis is presented in the accompanying technical memorandum.

The LWG feels that it is important to note that the approach presented in the draft BERA for evaluating benthic risk is consistent with prior instruction from EPA and that the apparent discrepancies reflect changes to EPA's preferences. In particular:

- EPA has reversed its preference for normalizing bioassay mortality data at least three times dating back to 2005.
    - The February 24, 2005, EPA-approved benthic approach stipulated that normalized bioassay mortality responses should be calculated as the difference between test and control mortality (test − control) (EPA 2005a). This is the method that was used in the draft BERA.



6650 SW Redwood Lane, Suite 333, Portland OR 97224

- A directive letter from EPA dated October 26, 2005, stipulated that normalized bioassay mortality responses should be calculated as the quotient of test and control mortality (test/control) (EPA 2005b).

- A December 22, 2005, letter from EPA referred back to the February 24, 2005, benthic approach and stated that the normalization should be test − control (EPA 2008a). The LWG notes that both the difference method and the quotient method are standard methods and believes that either method is reasonable but that switching methods creates unnecessary rework and confusion and creates the impression of substantive discrepancies where none exist.

- In the *Portland Harbor Superfund Site Ecological Risk Assessment Interpretive Report: Estimating Risks to Benthic Organisms Using Predictive Models Based on Sediment Toxicity Tests* (Windward et al. 2006), the normalization calculation was performed according to the Washington State Department of Ecology SedQual/Environmental Information Management (EIM) approach (T/C for mortality and (C−T)/C for growth).

- In EPA's 2008 BERA problem formulation, the normalization calculation was not specified, so the test-control method was used per the most recent direction from EPA (EPA 2008b).

- EPA's July 17, 2009, communication stipulated that normalized bioassay mortality responses should be calculated as test/control (EPA 2009b).

- EPA's preference for defining bioassay hits has changed several times since EPA approved the benthic approach in its February 24, 2005, letter.

  - The February 2005 EPA-approved benthic approach proposed using the freshwater Regional Sediment Evaluation Team criteria (T−C > 10% and T−C > 25% for both mortality endpoints, T/C < 0.75 and T/C < 0.6 for *Hyalella* growth, and T/C < 0.8 and T/C < 0.7 for *Chironomus* growth) (EPA 2005a).

  - The March 2006 benthic interpretive report (Windward et al. 2006) and the February 2007 Round 2 report (Integral et al. 2007) used EPA-directed hit definitions of 10, 20, and 30%.

  - EPA's February 2008 BERA problem formulation defined hits as follows: response between 90 and 80% of control = "minor effect" response between 80 and 70% of control = "moderate effect" response < 70% of control = "severe effect" (EPA 2008b).

  - On June 1, 2008, the remedial investigation data were locked down and the modeling effort for the BERA started.

  - On July 11, 2008, the LWG and EPA verbally agreed to the following hit definitions for the modeling effort: mortality increase > 20% relative to control, *Chironomus* biomass reduction > 30% relative to control and *Hyalella* biomass reduction > 40% relative to control (Toll 2008).

  - In September 2008, the LWG received the MacDonald and Landrum (2008) report, which proposed defining bioassay hits relative to reference conditions (the "reference envelope approach"). The LWG and EPA agreed to use the reference envelope approach. Neither the MacDonald and Landrum report nor the Calcasieu Estuary BERA cited by MacDonald and


6650 SW Redwood Lane, Suite 333, Portland OR 97224

Landrum fully defined the calculation procedures, but the conceptual details were complete, and the LWG implemented the reference envelope approach in a manner consistent with the information EPA had provided.

- Based on a request from EPA, the LWG explained the reference envelope approach used to define bioassay hits for the draft BERA and provided EPA the resultant hit thresholds during a June 18, 2009, teleconference. EPA expressed some concerns about the LWG's interpretation of the reference envelope approach and suggested that the LWG use an approach that would replicate the calculations presented in a table (Table E2-5) from the Calcasieu Estuary BERA. However, the documentation in the Calcasieu BERA was incomplete and internally inconsistent and the complete reference dataset was not provided in the document, so REVs could not be recreated based solely on the information provided. To determine the method it should replicate, the LWG solicited input from scientists involved in the Calcasieu BERA and then used a trial-and-error approach until the results in Table E2-5 of the Calcasieu BERA were matched as closely as possible. he LWG then verbally offered to revise the benthic BERA to conform to the Table E2-5 reference envelope calculation procedures. EPA did not respond to that offer, so the reference envelope method for defining hit thresholds was not revised.

- EPA's July 17 and 31, 2009, communications defined new reference envelope calculation procedures for defining bioassay hits (EPA 2009b, c). These new procedures are inconsistent with EPA's standing directives regarding bioassay interpretation.

The LWG does not agree that the procedures stipulated in EPA's July 17 and 31, 2009, communications are technically necessary or appropriate or that they substantively change the benthic interpretation in the BERA. The LWG would have preferred to eliminate the discrepancies between EPA's preferred procedures and the draft BERA prior to document submittal; but per EPA's direction (EPA 2009c), we submitted the document without reconciling the discrepancies.

Our team subsequently engaged in an effort to reconcile the differences between the draft BERA procedures and EPA's June, July, and August 2009 advice and instructions. Specifically, LWG and EPA ecological risk assessors met on August 26, 2009, to satisfy "check-in" terms of EPA's July 31, 2009, instructions (EPA 2009a). EPA's satisfaction regarding the check-in terms was confirmed in writing by EPA on August 28, 2009 (EPA 2009c), enabling us to proceed with the work presented in the accompanying benthic reanalysis technical memorandum.

The benthic reanalysis technical memorandum is an important step toward reconciling the differences between the draft BERA procedures and EPA's June, July, and August 2009 advice and instructions. We hope that it will help facilitate your timely review of the draft BERA.

Sincerely,

Bob Wyatt

Copies: LWG Management
LWG Repository



6650 SW Redwood Lane, Suite 333, Portland OR 97224

**References**

EPA. 2005a. EPA letter dated February 24, 2005, to Lower Willamette Group (from E. Blischke and C. Humphrey to J. McKenna and R. Wyatt) regarding Portland Harbor RI/FS: Estimating risks to benthic organisms using sediment toxicity tests (approval letter). US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2005b. Memo to Lower Willamette Group. Subject: benthic interpretive approach for Portland Harbor Ecological Risk Assessment. October 26, 2005. US Environmental Protection Agency, Portland, OR.

EPA. 2008a. EPA e-mail dated July 14, 2008 (Eric Blischke to John Toll), regarding bioassay evaluation. Remedial Project Manager, US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2008b. Problem formulation for the Baseline Ecological Risk Assessment at the Portland Harbor Site. Report and letter dated February 15, 2008 to Lower Willamette Group (from E. Blischke and C. Humphrey to J. McKenna and R. Wyatt). US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2009a. EPA e-mail dated August 28, 2009, to Lower Willamette Group (from B. Shephard to J. Toll) regarding record of decisions from 8/26/09 meeting to define EPA-approved bioassay reference envelope values. 8/28/09. US Environmental Protection Agency Region 10 Headquarters, Seattle, WA.

EPA. 2009b. EPA e-mail dated July 17, 2009, to Lower Willamette Group (from E. Blischke to R. Wyatt) regarding Portland Harbor RI/FS: interpretation of sediment bioassay results. US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2009c. EPA e-mail dated July 31, 2009, to Lower Willamette Group (from E. Blischke to J. McKenna) regarding Portland Harbor RI/FS: benthic interpretation. US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

Integral, Windward, Kennedy/Jenks, Anchor. 2007. Portland Harbor RI/FS: Comprehensive round 2 site characterization summary and data gaps analysis report, plus addenda. IC07-0004. Prepared for Lower Willamette Group. Integral Consulting, Inc., Mercer Island, WA; Windward Environmental LLC, Seattle, WA; Kennedy/Jenks Consultants, Portland, OR; Anchor Environmental, LLC, Portland, OR.

MacDonald DD, Landrum PF. 2008. An evaluation of the approach for assessing risks to the benthic invertebrate community at the Portland Harbor Superfund site. Preliminary draft. Prepared for US Environmental Protection Agency. September 2008. MacDonald Environmental Sciences, Ltd., Nanaimo, BC, and Landrum and Associates, Ann Arbor, MI.

Toll J. 2008. E-mail dated June 12, 2008, to EPA (Eric Blischke) regarding bioassay interpretation and benthic modeling "agreement". Windward Environmental, Seattle, WA.

Windward. 2009. Portland Harbor RI, Appendix G: Baseline ecological risk assessment. Draft. Prepared for the Lower Willamette Group. August 19, 2009. Windward Environmental LLC, Seattle, WA.

Windward, Avocet, TerraStat. 2006. Portland Harbor Superfund Site ecological risk assessment. Interpretive report: estimating risks to benthic organisms using predictive models based on sediment toxicity tests. Draft. WE-06-0002. Prepared for Lower Willamette Group. Windward Environmental LLC, Seattle, WA; Avocet Consulting, Kenmore, WA; TerraStat Consulting Group, Snohomish, WA.

LWG
LOWER WILLAMETTE GROUP

PORTLAND HARBOR RI/FS

# BENTHIC TOXICITY REANALYSIS TECHNICAL MEMORANDUM

**DRAFT**

| |
|---|
| **DO NOT QUOTE OR CITE**<br>This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part. |

November 13, 2009

**Prepared for**
The Lower Willamette Group

**Prepared by**
Windward Environmental LLC

WE-09-0004

RECOMMENDED FOR INCLUSION IN ADMINISTRATIVE RECORD

# TABLE OF CONTENTS



# LIST OF APPENDICES

# LIST OF TABLES

# LIST OF FIGURES



# LIST OF MAPS

# 1.0 INTRODUCTION

This memo provides a follow-up to the Portland Harbor draft baseline ecological risk assessment (BERA) (Windward 2009), in which risks to the benthic invertebrate community from exposure to sediment-borne contaminants in the Study Area were assessed by the Lower Willamette Group (LWG). The follow-up is provided in order to document the effects of specific changes to the method used in the draft BERA for calculating reference envelope values (REVs). REVs are estimates of the low end of the range of survival and biomass measurements in the upstream reach of the Lower Willamette River. They're used in the draft BERA to help define bioassay hit thresholds, at EPA's direction (EPA 2008). The memo describes and discusses statistical and technical issues involved in the execution of these methods and the interpretation of results. It also describes the resulting effects on designations of toxicity, values of site-specific sediment-quality guidelines (SQGs), success of SQGs in predicting toxicity at the site, and the locations of potential benthic community risk.

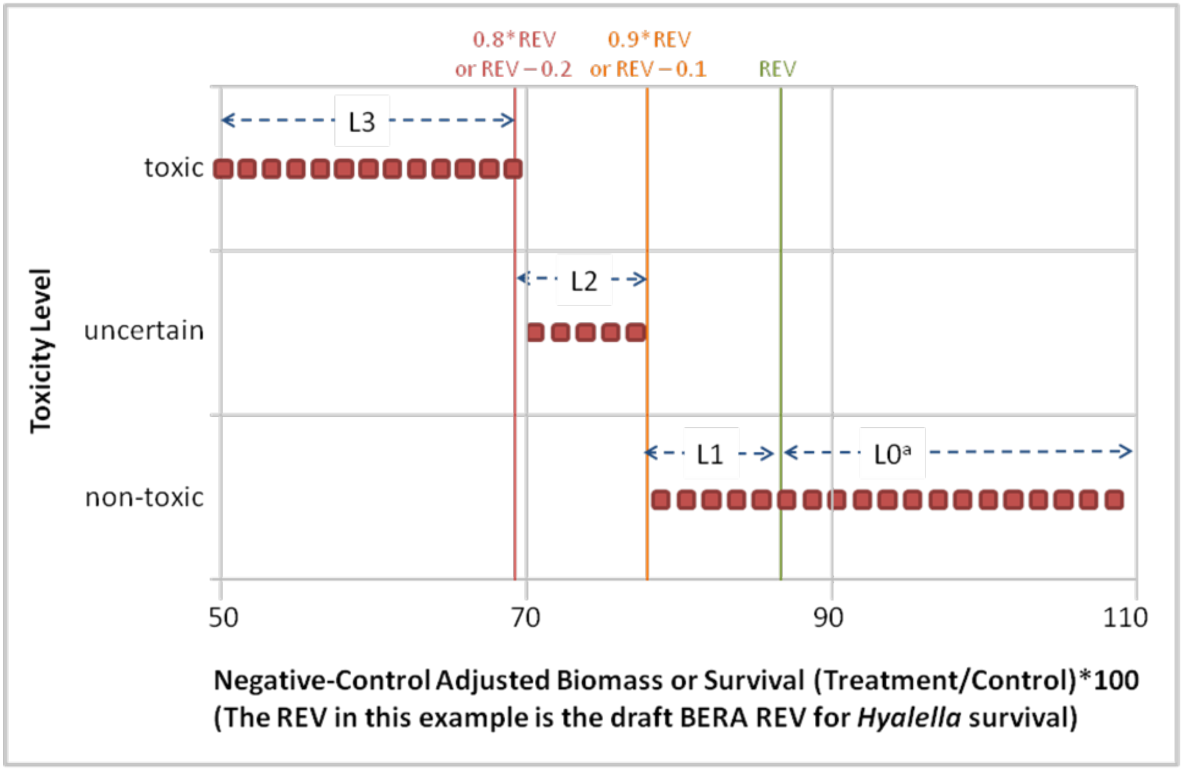
The work presented in this memo helps quantify the uncertainty in predicting toxicity to benthic organisms and risk to the benthic community exposed to Study Area sediments. The results indicate an encouraging degree of consistency, both across REV calculation methods and across bioassay endpoints. SQG exceedance areas varied somewhat by method of deriving hit thresholds (165 acres by the draft BERA method, with an uncertainty range of 121 to 193 acres based on the methods used in this memo). The level of consistency should provide confidence that robust risk management decisions may be made despite uncertainty regarding precise bioassay hit thresholds. The results of these investigations lead the LWG to recommend that the conclusions of the draft BERA be carried forward into the Feasibility Study because they represent a central tendency in the range of uncertainty about benthic toxicity and benthic community risk.

## 1.1 DEVELOPMENT OF THE DRAFT BERA

The draft BERA was compiled under oversight by the US Environmental Protection Agency (EPA) and was submitted for review on September 2, 2009. The approach to assess benthic community risks was developed and refined over a number of years based on EPA's formal comments on LWG project deliverables, as well as workshops, conference calls, e-mails, and letters. In 2008, EPA brought in an outside party to provide a peer review of the benthic assessment approach (MacDonald and Landrum 2008), which recommended the use a reference envelope approach to evaluate sediment toxicity to benthic organisms rather than a fixed difference from control responses that had been employed in previous assessments. The peer review described many details of a reference envelope approach and cited the reference envelope approach used in the draft BERA of a remedial investigation and feasibility study (RI/FS) for the Calcasieu Estuary (MacDonald Environmental 2002). EPA adopted the peer review recommendation to use a reference envelope approach and issued a directive for the LWG to implement this in the draft BERA (EPA 2008).

To implement the reference envelope approach, the LWG and EPA first agreed on a set of upriver bioassay data and bioassay data from two locations near the upper end of the Study Area that could be used to characterize background conditions in the Lower Willamette

River (EPA 2009c). At each of these sites, the level of negative control-adjusted growth and survival of *Chironomus dilutus* and *Hyalella azteca* was calculated and, for each test endpoint a REV was calculated. If the negative control-adjusted bioassay response at a Study Area location exceeded 90% of the REV[1], the station was classified as non-toxic for that endpoint and designated as Level 0 (if the response was >REV or not significantly different from negative control) or Level 1. If the response fell below 80% of the REV[2], the station was classified as toxic and designated Level 3. If the response fell between 80 and 90% of the REV, the station was classified as toxicity uncertain and designated Level 2 (Figure 1).




[a] Samples also were classified L0 if the mean test and negative control responses weren't statistically significantly different (one-sided t-test, alpha = 0.05).

**Figure 1-1. Schematic showing relationship between REV, toxicity levels, and range of negative control adjusted bioassay responses**

The methods for calculating REVs and toxicity levels in the draft BERA were taken from the peer review and Appendix E2 of the Calcasieu BERA. Distribution fitting software was

[1] The draft BERA used REV-0.1.

[2] The draft BERA used REV-0.2.

used to find the statistical distribution that best fit the reference station toxicity data for each bioassay type and to compute the fifth percentile of that distribution.

## 1.2 METHODS INVESTIGATIGATED IN THIS MEMO

EPA confirmed that the reference envelope approach as implemented by LWG in the draft BERA (Windward 2009) was generally consistent with its 2008 directive (EPA 2009c) but raised concerns during a June 18, 2009, conference call regarding several details of the calculation procedures. It was suggested in that conference call that the reference envelope approach for the draft BERA should be consistent with the methods used to obtain results presented in Table E2-5 of the BERA for the Calcasieu Estuary (MacDonald Environmental 2002), hereafter referred to as the Calcasieu BERA. After studying the Calcasieu BERA, it was determined that the major differences between the draft BERA and the Calcasieu BERA involved methods for:

- Normalizing survival data (the Calcasieu BERA used treatment survival/negative control survival; the draft BERA used treatment survival - negative control survival), consistent with regional guidance (RSET 2009) and past project direction from EPA (Windward et al. 2005)
- Computing the reference value (the Calcasieu BERA used the lower prediction limit of a two-sided prediction interval calculated using log-transformed reference data for each endpoint as opposed to the draft BERA's use of the 5$^{th}$ percentile value of the theoretical probability distribution that best fit the reference data for each endpoint)
- Computing risk levels greater than reference risk (the Calcasieu BERA added 10 and 20% of the REV to the REV to compute Level 2 and 3 hit thresholds; the draft BERA added constant values of 10 and 20% to the REV)

On July 17, 2009, EPA followed up on the June 18, 2009, conference call with written instructions, which were refined again on July 31, 2009 (EPA 2009a, b). EPA's July instructions differed from the approach used in the Calcasieu BERA (MacDonald Environmental 2002) but closely resembled the procedures used in the draft BERA. The July instructions differed from the June instructions in requiring that the REV be calculated as the 5$^{th}$ percentile of a best-fitting distribution with the stipulation that EPA approve the choice of best-fitting distribution for each endpoint. EPA agreed that these most recent instructions did not have to be incorporated into the draft BERA (Windward 2009) because of the potential impact on the RI/FS schedule (EPA 2009d).

EPA's July 17 and 31, 2009, instructions (EPA 2009a, b) were further modified during a meeting between LWG and EPA ecological risk assessors on August 26, 2009, to satisfy "check-in" terms of the July 31, 2009, instructions (Toll 2009a). The modifications and EPA's satisfaction regarding the check-in terms were confirmed in writing by EPA on August 28, 2009 (Toll 2009b).

The advice and instructions provided by EPA in June, July, and August 2009 all relate to the details of implementing a reference envelope approach – in particular, the details of methods used to calculate REVs and hit thresholds from raw bioassay data. Differences in methods produce changes in hit thresholds, the hit classifications assigned to the bioassay stations based on these hit thresholds, the resulting site-specific sediment quality guidelines (SQGs), and the performance of SQGs (generic[3] or site-specific) as predictors of toxicity.

This memorandum presents the analyses that were performed, based on EPA's June, July, and August 2009 advice and instructions, to classify bioassay hits, recalculate site-specific SQGs, evaluate the reliability of SQG models and reassess potential benthic risk areas (PBRAs). It includes both the Calcasieu BERA methods that EPA recommended in June 2009 and the final methods EPA instructed the LWG to use in July and August 2009. Investigating two different methods that were both considered appropriate by EPA allowed us to quantify the uncertainty in REVs.

The memorandum examines whether and to what extent the overall conclusions of the draft BERA (Windward 2009) regarding benthic community risk are sensitive to the details of pertaining to the implementation of the reference envelope approach, as represented by the procedures applied in the draft BERA and the procedures articulated in the advice and instructions provided by EPA in the summer of 2009, and to what extent any conclusions in the BERA should be modified based on these new analyses.

The details of the memorandum are presented as follows:

- Section 2.0 describes the alternative procedures for determining REVs and subsequent hit thresholds based on EPA 2009 instructions.
- Section 3.0 presents the sediment toxicity results based on the alternative hit thresholds and the draft BERA (Windward 2009) hit thresholds.
- Section 4.0 presents the resulting alternative site-specific SQGs derived from the benthic predictive models and discusses their reliability at predicting bioassay hits.
- Section 5.0 presents the reliability of generic SQGs at predicting bioassay hits.
- Section 6.0 presents a comparison of the three approaches to assess benthic risk, hereafter referred to as the draft BERA, Calcasieu, and EPA 2009 approaches).

[3] Generic SQGs are non-site-specific values drawn from the literature. Typically, they are based on multiple-study, multiple region data or on a consensus among different sets of published SQGs.

# 2.0 ALTERNATIVE REFERENCE ENVELOPE AND HIT CLASSIFICATION PROCEDURES

This section investigates two alternative procedures for calculating REVs and compares the results to REVs calculated for the draft BERA (Windward 2009). The first set of procedures is derived from the Calcasieu BERA (MacDonald Environmental 2002). The second set is based on the instructions EPA provided in July 17 and 31, 2009, e-mails to the LWG, and refined during the August 26, 2009, meeting between EPA and LWG ecological risk assessors.

## 2.1 CALCASIEU BERA PROCEDURES

MacDonald and Landrum (2008) provided a general description of a reference envelope approach and cited Appendix E2 of the Calcasieu BERA (MacDonald Environmental 2002) as the source of the reference envelope calculation methods. During the June 18, 2009, teleconference, EPA expressed concern that the LWG had not implemented the Calcasieu approach correctly. EPA acknowledged that the LWG had implemented the methods as described in the methods section of the Calcasieu BERA (Appendix E2) but pointed out some (though not all) of the inconsistencies between what was described in the methods section of Appendix E2 and what was presented in the results section of that appendix. EPA identified Table E2-5 as what it considered to be the correct example of the Calcasieu reference envelope calculation methods.

A close review of the Calcasieu BERA revealed that EPA's understanding of the Calcasieu approach differed from the description in the document and that there appeared to be discrepancies both within the Calcasieu BERA and between the Calcasieu BERA and the MacDonald and Landrum report (2008) as to how to calculate reference envelope values. For example, the introduction of the Calcasieu BERA described the REV as a 95$^{th}$ percentile confidence interval, a 2.5$^{th}$ percentile, the "lower limit of the normal range (i.e., 95% LCL)" and as a lower 95% prediction limit (rather than a percentile or confidence limit).

The table in the Calcasieu BERA that displays the actual reference thresholds (Table E2-5) appears to follow the approach described in the risk characterization section in that it used a lower prediction limit (LPL). Based on our attempts to duplicate the values, it appears that the REV was calculated as a two-sided 95% LPL. It is unclear from the Calcasieu BERA whether or not the toxicity data were subjected to any transformations prior to the calculation of the REV.

Because the documentation in the Calcasieu BERA was incomplete and internally inconsistent and the complete reference dataset was not provided in the document, REVs could not be recreated based solely on the information provided. To determine the method it should duplicate, the LWG solicited input from scientists involved in the Calcasieu BERA and then used a trial-and-error approach until the results in Table E2-5 of the Calcasieu BERA were matched as closely as possible.

The LWG considers the Calcasieu procedures described in Table 2-1 to be an accurate representation of the methods used in the Calcasieu BERA that EPA recommended during the June 18, 2009, reference envelope teleconference. Further details on the Calcasieu procedures are presented in Appendix A.

**Table 2-1. Summary and Comparison of Alternative Reference Envelope Procedures**

| Risk Assessment Step | Calcasieu | Draft BERA | EPA 2009 |
|---|---|---|---|
| Treatment of sample and reference duplicates | Averaged duplicates | Retained as individual samples and used most conservative hit designation of replicates | Averaged duplicates |
| Mortality endpoint | Expressed as % survivors | Expressed as % mortality | Expressed as % survivors |
| Negative control normalization | Calculated as ratio of test/ negative control | Calculated as difference of test minus negative control | Calculated as ratio of test/ negative control |
| Reference envelope calculations | Calculated as back-transformed lower (2.5%) prediction limit of log-transformed negative control-adjusted bioassay data | Calculated as lower 5$^{th}$ percentile for biomass/upper 95$^{th}$ percentile for mortality using best-fit distribution, which varied by endpoint[a] | Used best-fit distribution for a given endpoint (survival or biomass); calculated lower 5$^{th}$ percentile of distribution that best fit lower tail, following EPA approval of distribution[b] |
| Data transformation | Log-transformed data; then back-transformed data | No transformation | No transformation per August clarification |

Sources: MacDonald Environmental (2002), Windward (2009), EPA (2009a, b)

a The BERA used a lognormal distribution as the best fit for *Chironomus* mortality, the Weibull distribution for *Chironomus* biomass, and a log-logistic distribution for both *Hyalella* mortality and biomass.

b EPA selected the logistic distribution for *Chironomus* survival, the Weibull distribution for *Chironomus* biomass, a beta distribution for *Hyalella* survival, and an exponential distribution for *Hyalella* biomass.

BERA – baseline ecological risk assessment

EPA – US Environmental Protection Agency

## 2.2 EPA PROCEDURES

In its July 17 and 31, 2009, correspondences with the LWG (EPA 2009a, b), EPA instructed that the best-fit distribution of the reference area data be used to calculate a lower 5$^{th}$ percentile to define the reference threshold, that the selected best-fit distribution be approved by EPA before the calculation of reference envelope values, and that data be log-transformed prior to the selection of the best-fit distribution (no transformations were required or used in the draft BERA). When EPA and the LWG met on August 26, 2009, to select the final distributions to be used, EPA agreed with the LWG that data transformation was unnecessary because the statistical distribution fit provided by the @RISK distribution-fitting software addresses the issue of data skewness, which was the source of EPA's data transformation concern.

The Calcasieu, EPA 2009, and draft BERA reference envelope procedures collectively represent a range of interpretations of the reference envelope approach that EPA directed the LWG to use (EPA 2009b). A summary and comparison of the individual elements of each way of implementing the reference envelope approach is provided in Table 2-1. Additional details on the procedures for deriving the reference envelope values are presented in Appendix A.

The final reference envelope values derived from each method are presented in Table 2-2.

**Table 2-2. Reference Envelope Values Based on the Calcasieu, Draft BERA, and EPA 2009 Procedures**

| | Calcasieu | Draft BERA | EPA 2009 |
|---|---|---|---|
| Test and Endpoint | Reference Value (%)[a] | Reference Value (%)[a, b] | Reference Value (%)[a] |
| *Chironomus dilutus* survival | 89.8 | 91.9 | 93.9 |
| *Chironomus dilutus* biomass | 83.4 | 88.7 | 91.0 |
| *Hyalella azteca* survival | 87.2 | 86.7 | 88.1 |
| *Hyalella azteca* biomass | 64.7 | 67.1 | 73.6 |

Sources: MacDonald Environmental (2002), Windward (2009), EPA (2009a, b)

[a] 5th percentile of negative control-adjusted (test/ negative control) survival and biomass endpoints for reference stations. For draft BERA survival, negative control adjustment was calculated as (100 - (test mortality - negative control mortality)).

[b] The values have been expressed as survival and biomass to be comparable with the Calcasieu and EPA 2009 thresholds.

BERA – baseline ecological risk assessment

EPA – US Environmental Protection Agency

## 2.3 ALTERNATIVE TOXICITY THRESHOLDS AND HIT CLASSIFICATION SCHEME

In the draft BERA (Windward 2009), the toxicity test results were statistically compared to negative control results and numerically compared to four effects thresholds to define the potential for toxicity. Effects levels were based on incremental differences from REVs (see Section 2.2 for a description of how REVs were calculated) and were defined in the draft BERA as:

- **Level 0 –** mean response not significantly different from negative control[4] mean or mean negative control-adjusted response ≥ REV
- **Level 1 –** mean response significantly different from the negative control[2] mean and REV > mean negative control-adjusted response ≥ REV – 10%

[4] Using one-sided t-test, alpha = 0.05.

- **Level 2 –** mean response significantly different from the negative control[2] mean and REV – 10% > mean negative control-adjusted response ≥ REV – 20%
- **Level 3 –** mean response significantly different from the negative control[2] mean and REV – 20% ≥ mean negative control-adjusted response

The toxicity thresholds developed for the draft BERA (Windward 2009) were based on absolute differences (10 and 20%) from REVs to define hit thresholds. Toxicity thresholds derived from the Calcasieu and EPA 2009 methods are based on relative difference from REVs:

- **Level 0 –** mean response not significantly different from the negative control[5] mean or mean negative control-adjusted response ≥ REV
- **Level 1 –** mean response significantly different from the negative control[2] mean and REV > mean negative control-adjusted response ≥ 0.9*REV
- **Level 2 –** mean response significantly different from the negative control[2] mean and 0.9*REV > mean negative control-adjusted response ≥ 0.8*REV
- **Level 3 –** mean response significantly different from the negative control[2] mean and 0.8*REV > mean negative control-adjusted response

Table 2-3 presents the toxicity thresholds that were used to classify the magnitude of toxicity, based on the Calcasieu, draft BERA, and EPA 2009 procedures.

[5] Using one-sided t-test, alpha = 0.05.

**Table 2-3. Toxicity Thresholds Based on the Calcasieu, Draft BERA, and EPA 2009 Procedures**

| | Calcasieu | | | Draft BERA | | | EPA 2009 | | |
|---|---|---|---|---|---|---|---|---|---|
| **Test and Endpoint** | **REV (L1) Threshold (%)[a]** | **Low (L2) Threshold (%)[b]** | **High (L3) Threshold (%)[c]** | **REV (L1) Threshold (%)[a, d]** | **Low (L2) Threshold (%)[b, d]** | **High (L3) Threshold (%)[c, d]** | **REV (L1) Threshold (%)[a]** | **Low (L2) Threshold (%)[b]** | **High (L3) Threshold (%)[c]** |
| *Chironomus dilutus* survival | 89.8 | 80.8 | 71.8 | 91.9 | 81.9 | 71.9 | 93.9 | 84.5 | 75.1 |
| *Chironomus dilutus* biomass | 83.4 | 75.1 | 66.7 | 88.7 | 78.7 | 68.7 | 91.0 | 81.9 | 72.8 |
| *Hyalella azteca* survival | 87.2 | 78.5 | 69.8 | 86.7 | 76.7 | 66.7 | 88.1 | 79.3 | 70.5 |
| *Hyalella azteca* biomass | 64.7 | 58.2 | 51.7 | 67.1 | 57.1 | 47.1 | 73.6 | 66.2 | 58.9 |

Sources: MacDonald Environmental (2002), Windward (2009), EPA (2009a, b)

[a] From Table 2-2.

[b] The negative control-adjusted survival and biomass endpoints must be less than the corresponding low threshold (0.9*REV for alternative methods; REV-10% for the draft BERA), and the mean test response must be statistically less than the mean negative control response using a one-tailed t-test ($p < 0.05$) for the sediment to be considered as having an adverse effect on benthic invertebrates.

[c] The negative control-adjusted survival and biomass endpoints must be less than the corresponding high threshold (0.8*REV for alternative methods and REV-20% for the draft BERA), and the mean test response must be statistically less than the mean negative control response using a one-tailed t- test ($p < 0.05$) for the sediment to be considered the sediment as having an adverse effect on benthic invertebrates.

[d] Values have been expressed as survival and biomass to be comparable with the Calcasieu and EPA 2009 thresholds.

BERA – baseline ecological risk assessment

EPA – US Environmental Protection Agency

L1 – Level 1

L2 – Level 2

L3 – Level 3

REV – reference envelope value

# 3.0 HIT CLASSIFICATION RESULTS

The toxicity test data used in the draft BERA (Windward 2009) to assess benthic community risks included the 10-day sediment test, which measured survival and biomass in *Chironomus dilutus*, and the 28-day sediment test, which measured survival and biomass in *Hyalella azteca*. The toxicity thresholds, based on the REVs, were used to classify the bioassay results as Levels 0, 1, 2, or 3. Table 3-1 summarizes the hit classification results for the 269 bioassay stations in the Study Area. More than half of the toxicity test results (55 to 87% depending on endpoint and reference envelope method) were classified as Level 0, 4 to 18% (depending on endpoint and method) were classified as Level 1, 1 to 13% were classified as Level 2, and 6 to 16% were classified as Level 3.

Across all endpoints and methods for classifying the magnitude of toxicity in the 269 samples tested, 121 sample classifications agreed and 148 differed. There was a fair amount of noise among Level 0 and Level 1 classifications but both of these classes are considered non-toxic. At Levels 2 and 3 in particular there was a high degree of concordance, and differences, when they did occur, were due primarily to the *Hyalella* biomass endpoint. *Hyalella* survival results were the most consistent in the classification of the Level 2 and Level 3 hits across all reference envelope methods (only two samples differed in classification among methods).

*Chironomus* survival in the reference samples ranged from 91.2 to 113.2% of the negative control response. Mean *Chironomus* negative control survival varied by 4% (from 85% to 88.75%) in the three negative control batches compared to reference samples and from 85% to 95% in all negative control batches, indicating that uncertainty associated with the hit classifications resulting from control batch effects is low for this endpoint.

*Chironomus* biomass in the reference area test samples ranged from 82.3 to 136.0% of the negative control response, and *Chironomus* negative control biomass varied by 20% (from 0.72 mg to 0.89 mg) in the reference samples. Chironomus negative control biomass ranged from 0.69 mg to 1.3 mg across all 14 batches used to test study area samples.

**Table 3-1. Study Area Toxicity Data Compared to the Negative Control and the Calcasieu, Draft BERA, and EPA 2009 Toxicity Thresholds**

| | Number of Sampling Locations | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Chironomus* Survival | | | *Hyalella* Survival | | | *Chironomus* Biomass[a] | | | *Hyalella* Biomass[a] | | |
| **Category** | **Calcasieu** | **Draft BERA** | **EPA 2009** | **Calcasieu** | **Draft BERA** | **EPA 2009** | **Calcasieu** | **Draft BERA** | **EPA 2009** | **Calcasieu** | **Draft BERA** | **EPA 2009** |
| Level 0: Not significantly different ($p \geq 0.05$) from negative control or significantly different but less than REV | 219 | 219 | 213 | 234 | 235 | 231 | 215 | 201 | 197 | 216 | 190 | 148 |
| Level 1: Significantly different ($p < 0.05$) from negative control and ≥ low toxicity threshold | 13 | 14 | 12 | 14 | 13 | 17 | 10 | 22 | 20 | 24 | 56 | 49 |
| Level 2: Significantly different ($p < 0.05$) from negative control and< low toxicity threshold and ≥ high toxicity threshold | 9 | 11 | 12 | 3 | 4 | 2 | 8 | 9 | 10 | 13 | 8 | 34 |
| Level 3: Significantly different ($p < 0.05$) from negative control and < high toxicity threshold | 28 | 25 | 32 | 18 | 17 | 19 | 36 | 37 | 42 | 16 | 15 | 38 |

Sources: MacDonald Environmental (2002), Windward (2009), EPA (2009a, b)

[a] The biomass endpoint was defined as the total mass of survivors in a sample.

BERA – baseline ecological risk assessment

EPA – Environmental Protection Agency

REV – reference envelope value

The *Hyalella* survival response in the reference samples ranged from 85.0 to 102.6% of the negative control response, and the *Hyalella* negative control survival varied by 10% (from 90% to 100%) in the reference samples.

The *Hyalella* biomass response in the reference samples ranged from 74.4 to 113.2% of the negative control response, and the *Hyalella* negative control biomass varied by >100% (from 0.17 mg to 0.38 mg) in the reference samples and from 89% to 100% in all control batches. The variability in negative control biomass of *Hyalella* is much higher than for all other endpoints. It is not clear whether the sources of this variance are random or are created by non-random laboratory conditions that might also apply to paired test data but caution should be used in interpreting the *Hyalella* biomass hit classifications, particularly when small differences from negative control are considered indications of toxicity, as with the EPA 2009 method, because control-adjusted toxicity classifications based on this endpoint are poorly predicted by sediment chemistry data (Appendix B).

For example, replicate variability in test and control responses is high enough that the 95% confidence intervals on many station mean control-normalized responses include a range of responses that cover more than one level of toxicity. In many cases the lower 95% confidence bound on the response is lower than the Level 3 hit threshold while the upper 95% confidence bound is greater than the Level 3 or even the Level 2 or Level 1 thresholds, indicating that the station's bioassay hit classification is uncertain due to within station replicate variability.

EPA selected an exponential distribution as the best fit for representing the *Hyalella* biomass reference data. The exponential distribution is an unusual distribution for bioassay response data from a reference area (the data would be expected to have a two-tailed distribution) and was an artifact of the particularly low biomass in negative control samples from two test batches.

The *Hyalella* toxicity tests were conducted in 14 batches in 2004 and 2007, and the reference samples were tested in three separate batches. The supplier of the *Hyalella* used in the tests raised the amphipods in 20 separate cultures (the cultures were mixed once or twice a year). The large numbers of batches and cultures introduce variability into the biomass data based on the different *Hyalella* cultures used in the tests and potential seasonality or variation in the health of the cultures. Inherent variability in cultures and the health of the cultures over time could be confounding the hit classification results for this endpoint. The culture effect on mean *Hyalella* biomass has not been quantified, even in ASTM round-robin testing, but it could be affecting bioassay comparisons to the REV. In the ASTM round-robin testing (ASTM 2007), a relatively high variability was observed in *Hyalella* biomass data generated by eight laboratories using eight different in-house cultures based on MDDs of about 24 and 40% of the mean negative control response.

These factors indicate that the *Hyalella* biomass hit classifications contain errors, most of which are not controlled by data normalization. This, combined with an inability to reliably predict hits from sediment chemistry, indicates that the *Hyalella* biomass endpoint is an unreliable indicator of benthic toxicity, at least when interpreted using the EPA 2009 hit thresholds presented in Table 2-3.

Study Area bioassay hit classifications for the four test endpoints and three approaches (i.e., draft BERA, Calcasieu, and EPA) are presented on Maps 3-1 through 3-12. The maps show strong agreement across all bioassay endpoints and hit classification methods, except *Hyalella* biomass using the EPA 2009 method, which is inconsistent with all other endpoints and with the *Hyalella* biomass results by the draft BERA and Calcasieu methods.

Table 3-2 presents the pooled endpoint results for the three methods. With the pooled endpoint, the most conservative bioassay result determines the hit classification. The results of the pooled endpoint are similar for the three methods based on all four endpoints for the Calcasieu and draft BERA methods and three endpoints (excluding *Hyalella* biomass) for the EPA 2009 method. Pooled sample classifications resulted in 14% (Calcasieu and draft BERA approach) to 16% (EPA 2009 approach) of the total number of samples exceeding the Level 3 threshold. About 20 to 22% of the stations were classified as either Level 2 or Level 3 based on the pooled endpoint.

**Table 3-2. Pooled Toxicity Data Classifications for the Study Area Stations Using the Calcasieu, Draft BERA, and EPA 2009 Approaches**

| | Number of Stations Based on the Pooled Results | | |
|---|---|---|---|
| **Category** | **Calcasieu** | **Draft BERA** | **EPA 2009** |
| Level 0: Not significantly different ($p \geq 0.05$) from negative control or significantly different but less than REV | 168 | 149 | 169 |
| Level 1: Significantly different ($p < 0.05$) from negative control and ≥ low toxicity threshold | 43 | 65 | 41 |
| Level 2: Significantly different ($p < 0.05$) from negative control and < low toxicity threshold and ≥ high toxicity threshold | 20 | 18 | 15 |
| Level 3: Significantly different ($p < 0.05$) from negative control and < high toxicity threshold | 38 | 37 | 44 |

Sources: MacDonald Environmental (2002), Windward (2009), EPA (2009a, b)
BERA – baseline ecological risk assessment
EPA – Environmental Protection Agency
REV – reference envelope value

Map 3-13 and Map 3-14 display the areas of disagreement and agreement in toxicity sample classifications in Study Area. Of the 269 stations, 68% were classified as the same level by all methods. Most of the disagreement among methods was regarding Level 1 and Level 2 classifications. Disagreements regarding Level 3 hit classifications tended to occur in locations where other bioassays provided a preponderance of evidence for determining whether the area posed potential risk to the benthic community, and so the disagreements regarding hit classifications had little influence on the benthic community risk characterization.

The sediment toxicity assessment relies on the selection of a reference dataset and establishment of characteristic response values to represent conditions in the absence of site-related contaminants. EPA's selection of upstream stations applied additional criteria to limit the influence of any anthropogenic chemical inputs into the Lower Willamette River by requiring that candidate reference samples meet three different sets of sediment quality guidelines. This represents a conservative assessment of the range of toxicity responses that might be expected from sediments upstream of the Study Area, in which case all three methods would be biased to overestimate toxicity relative to anthropogenic background.

# 4.0 BENTHIC PREDICTIVE MODELS

As in the draft BERA (Windward 2009), the logistic regression model (LRM) and floating percentile model (FPM) were used to try to develop sets of site-specific SQGs that reliably predicted the bioassay results (using the EPA 2009 and Calcasieu hit classifications) from sediment chemistry data. In the draft BERA, the LRM failed to reliably predict the bioassay results, so it was not used to derive SQGs. The LRM has again failed to meet reliability criteria using the EPA 2009 and Calcasieu hit classifications and so, as before, was not used to derive SQGs. Appendix C presents a detailed description of the LRM and the reliability analysis outcomes. The remainder of this section describes:

- The bioassay and chemistry data used in the model (Section 4.1)
- The SQG derivation using the FPM with the Calcasieu and EPA 2009 toxicity thresholds (Section 4.2)

## 4.1 BIOASSAY AND CHEMISTRY DATA

As in the draft BERA (Windward 2009), predictive models were developed using the 293 samples with co-located sediment chemistry and toxicity test data. Of these surface sediment samples, 269 samples were collected within the Study Area, 2 samples were collected in the downtown reach, and 22 samples were collected in the upriver reach. Thirteen of the sampling locations in the Study Area were dredged after toxicity testing had been performed and are not part of the draft BERA risk dataset; however, these sampling locations were retained for use in model development. Separate models were developed for each of the four endpoints (i.e., *Chironomus* survival and biomass and *Hyalella* survival and biomass) and the pooled endpoint (defined by the highest hit classification at a particular station for any of the four individual endpoints).

The surface sediment chemistry dataset was the same as that used in the draft BERA (Windward 2009). Only detected values were used because non-detected chemistry values cannot be used to develop a predictive relationship between sediment chemistry and sediment toxicity (Avocet 2003). One exception was for non-detects that were part of a chemical group total (e.g., total PCBs). In these cases, one-half the detection limit was used in the group total. Chemical data qualified with N, NJ, and NJT[6] qualifiers were included in the models at EPA's direction (EPA 2006). Chemicals with fewer than 30 detected values were excluded from the modeling effort because at least 30 data points are needed to create a usable distribution for the development of SQGs based on the analysis of other datasets from Oregon and Washington (Avocet 2003). Certain conventional parameters (e.g., specific gravity, individual grain size, and total solids) were screened out of both models

[6] An N-qualifier signifies the presumptive evidence of an analyte; for metals, the matrix spike sample recovery was not within control limits, and for organics, the identification was tentative; the analyte exhibited low spectral match parameters but was present. A J-qualifier signifies an estimated value. A T-qualifier signifies that the value is an average or selected result (following standard project rules).

because they are not considered contaminants. However, other conventional parameters, including percent fines, bulk sediment ammonia, and sulfides, were retained in the two models because they enhanced the model performance. Percent fines was not considered a contaminant; and because ammonia and sulfides are natural sediment constituents that may cause toxicity, they were also not addressed as contaminants. Further details on the chemistry dataset are provided in the draft BERA, Section 6.2.1 (Windward 2009).

## 4.2 FLOATING PERCENTILE MODEL

The FPM uses a threshold approach to develop site-specific SQGs. The model allows the user to select a false negative rate (i.e., rate associated with the erroneous conclusion that the sediments are not toxic) and then adjusts the candidate SQG values until false positive rates (i.e., rates associated with the erroneous conclusion that sediments are toxic) are decreased to local minimums, given the false negative rate. The modeling was done using the automated FPM Microsoft Excel® spreadsheets provided by the Regional Sediment Evaluation Team (RSET) (Anderson 2008), and the designation of toxicity was based on the Calcasieu and EPA 2009 hit thresholds presented in Table 2-2. The FPM's candidate chemical list was refined using the same criteria and analysis of variance (ANOVA) evaluation as those in the draft BERA (Windward 2009). Table 4-1 presents the chemicals included in the FPM. Additional details on the FPM and the surface sediment chemistry dataset are presented in the draft BERA, Section 6.2.2.

**Table 4-1. Chemicals Retained in the Floating Percentile Model**

| Chemical | |
|---|---|
| **Metals** | |
| Antimony | Lead |
| Arsenic | Mercury |
| Cadmium | Silver |
| Chromium | Zinc |
| Copper | |
| **Butyltins** | |
| Dibutyltin | |
| **PAHs** | |
| Sum of 34 ESB PAHs | Total LPAHs[b] |
| Total benzofluoranthenes | Total PAHs[a, b] |
| Total HPAHs[a] | |
| **SVOCs** | |
| Benzyl alcohol | Carbazole |
| **Phthalates** | |
| Diethyl phthalate | |
| **Phenols** | |
| 4-Methyl phenol | Phenol |

**Table 4-1. Chemicals Retained in the Floating Percentile Model**

| Chemical | |
|---|---|
| **PCBs** | |
| Total PCBs | |
| **Pesticides** | |
| delta-HCH | Sum DDE |
| Dieldrin | Sum DDT |
| Endrin | Total chlordane |
| Endrin ketone | Total DDx |
| Sum DDD | |
| **TPH** | |
| PYO-PTO | |
| **Conventionals** | |
| Ammonia | Total percent fines |
| Sulfide | Total organic carbon |

[a] Includes benzo(a)anthracene, benzo(a)pyrene, benzo(b)fluoranthene, benzo(b+k)fluoranthene, benzo(g,h,i)perylene, benzo(j+k)fluoranthene, benzo(k)fluoranthene, chrysene, dibenzo(a,h)anthracene, fluoranthene, indeno(1,2,3-cd)pyrene, and pyrene.

[b] Includes 2-methylnaphthalene, acenaphthene, acenaphthene, acenaphthylene, anthracene, fluorine, naphthalene, and phenanthrene.

DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
DDT – dichlorodiphenyltrichloroethane
ESB – equilibrium partitioning sediment benchmark
HCH – hexachlorocyclohexane
HPAH – high-molecular-weight polycyclic aromatic hydrocarbon
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl
PYO – pyrogenic (petroleum compound)
PTO – petrogenic (petroleum compound)
SVOC – semivolatile organic compound
Total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT, and 4,4′-DDT)
TPH – total petroleum hydrocarbons

The FPM was run to derive low and high SQGs. As in the draft BERA (Windward 2009), the low SQGs were selected to reliably predict “clean” areas (i.e., areas that would be below Level 2 hit thresholds), and the high SQGs were selected to reliably predict “toxic” areas (i.e., areas that would exceed Level 3 hit thresholds).

SQGs were derived for both alternative sets of hit thresholds (Calcasieu and EPA 2009). Again, as in the draft BERA (Windward 2009), the final sets of SQGs from the FPM were developed after a series of model runs and subsequent reliability analyses. The reliability analyses were based on achieving the goals presented in the draft Washington State freshwater guidelines (Avocet 2003). Specifically, for the high SQGs, both false negative and false positive rates were required to be below 20%, and the overall reliability was required to be above 80%. In addition, predicted no-hit reliability was required to be above 90% in order to produce greater confidence in defining a sampling location as having no adverse effects (Avocet 2003).

Following the draft BERA methodology (Windward 2009), the requirement of < 20% false positive rates was suspended for the low SQGs because the narrative intent[7] for the low SQGs is to reliably predict “clean” areas. Allowing the false positive rate to exceed 20% simply means that low SQGs are conservatively biased to underestimate the total non-toxic area, which is appropriate for screening-level thresholds.

As described in Appendix B, each of the four endpoints (*Chironomus* and *Hyalella* survival and biomass) was run through the FPM for the low (Level 2) and high (Level 3) Calcasieu and EPA 2009 toxicity thresholds, producing individual endpoint models for four “method-level” combinations (i.e., Calcasieu-low, Calcasieu-high, EPA 2009-low and EPA 2009-high). From all of the individual endpoints in each method-level set that met the reliability criteria, the lowest SQG was selected for each chemical. These sets of “minimum SQGs” were then used to predict the conservative, pooled endpoint for that method-level (i.e., if any reliable individual endpoint indicated toxicity, the pooled endpoint was coded as a hit), and final error and reliability rates were computed for the pooled endpoint.

As was the case with the draft BERA hit thresholds (Windward 2009), the minimum SQGs for all four method-level combinations (i.e., Calcasieu-low, Calcasieu-high, EPA 2009-low and EPA 2009-high) failed to meet the established reliability criteria, so they were not used to define areas of unacceptable risk to the benthic community in the Study Area. Instead, after running the FPM to develop SQGs for each individual endpoint, the model was run again using the pooled endpoints (low and high). The same reliability criteria and methods were used for the pooled as for the individual endpoints. As in the draft BERA, the Calcasieu and EPA 2009 SQGs developed from the pooled response data met the model reliability criteria.

Table 4-2 presents the reliability results from FPM runs based on the pooled endpoints. The *Hyalella* biomass endpoint did not meet the acceptability criteria for the EPA 2009 hit thresholds (Appendix C), so *Hyalella* biomass was not included in the EPA 2009-low and EPA 2009-high pooled endpoints.

[7] The term “narrative intent” refers to the specific predictions associated with exposure to sediment chemical concentrations relative to SQGs. The concept of narrative intent is essential for understanding what the comparison of sediment chemistry data with any particular measurement endpoint means, in terms of the potential risk posed by that sediment to the benthic community. In general, the low SQGs are sediment chemical concentrations below which adverse effects on benthic invertebrates are not expected to occur, and the high SQGs are concentrations above which adverse effects on benthic invertebrates are somewhat likely to occur.

**Table 4-2. Overall Error and Reliability Rates for Initial FPM Based on the Pooled Endpoints**

| | Calcasieu | | EPA 2009 | |
|---|---|---|---|---|
| **Reliability Parameters** | **Low Threshold** | **High Threshold** | **Low Threshold** | **High Threshold** |
| % False negatives | 18.64 | 18.42 | 19.44 | 18.18 |
| % False positives | 31.62 | 10.59 | 52.97 | 18.07 |
| % Hit reliability | 81.36 | 81.58 | 80.56 | 81.82 |
| % No-hit reliability | 68.38 | 89.41 | 47.03 | 81.93 |
| % Predicted hit reliability | 39.34 | 53.45 | 47.03 | 44.44 |
| % Predicted no-hit reliability | 93.57 | 97.02 | 80.56 | 96.23 |
| % Overall reliability | 70.99 | 88.40 | 59.39 | 81.91 |

EPA – US Environmental Protection Agency
FPM – floating percentile model

The high-level site-specific SQGs for the Calcasieu and EPA 2009 approaches met the overall error and reliability criteria (Table 4-2) (Avocet 2003). The two sets of low SQGs did not achieve the desired false positive rate; but, as previously mentioned, because the narrative intent for this endpoint is to designate non-toxic areas, incorrectly predicting toxicity above this threshold was not considered as important as guaranteeing that false negative rates were maintained.[8] Table 4-3 presents the first sets of SQGs based on the pooled endpoints.

**Table 4-3. Initial Set of FPM SQGs**

| | Calcasieu | | EPA 2009 | |
|---|---|---|---|---|
| **Chemical** | **Low Toxicity Thresholds** | **High Toxicity Thresholds** | **Low Toxicity Thresholds** | **High Toxicity Thresholds** |
| **Metals (mg/kg dw)** | | | | |
| Antimony | 19.3 | 19.3 | 19.3 | 19.3 |
| Arsenic | 22.9 | 34 | 5.04 | 34 |
| Cadmium | 3.51 | 3.51 | 0.507 | 3.51 |
| Chromium | 224 | 224 | 224 | 224 |
| Copper | 562 | 562 | 493 | 562 |
| Lead | 1,290 | 1,290 | 1,290 | 1,290 |
| Mercury | 0.407 | 0.407 | 0.249 | 0.407 |

[8] Note that the high false positive rate for the low SQG set also has the effect of lowering the overall reliability rate below 80%.

**Table 4-3. Initial Set of FPM SQGs**

| | Calcasieu | | EPA 2009 | |
|---|---|---|---|---|
| **Chemical** | **Low Toxicity Thresholds** | **High Toxicity Thresholds** | **Low Toxicity Thresholds** | **High Toxicity Thresholds** |
| Silver | 1.72 | 1.72 | 0.412 | 1.72 |
| Zinc | 469 | 469 | 1,940 | 1,940 |
| **Butyltins** | | | | |
| Dibutyltin ion | 910 | 910 | 88.5 | 910 |
| **PAHs (µg/kg)** | | | | |
| Total benzofluoranthenes | 53,000 | 53,000 | 2,700 | 53,000 |
| Total HPAHs | 610,000 | 610,000 | 610,000 | 610,000 |
| Total LPAHs | 2,300 | 75,000 | 2,000 | 650,000 |
| Total PAHs | 1,300,000 | 1,300,000 | 330,000 | 330,000 |
| **SVOCs (µg/kg)** | | | | |
| Benzyl alcohol | 36 | 36 | 16 | 36 |
| Carbazole | 540 | 540 | 1,100 | 1,100 |
| **Phenols (µg/kg)** | | | | |
| 4-Methyl phenol | 96 | 125 | 96 | 96 |
| Phenol | 120 | 120 | 36 | 120 |
| **Phthalate (µg/kg)** | | | | |
| Diethyl phthalate | 370 | 370 | 7 | 370 |
| **PCBs** | | | | |
| Total PCBs (µg/kg) | 300 | 3,500 | 250 | 500 |
| **Pesticides (µg/kg)** | | | | |
| delta-Hexachlorocyclohexane | 0.912 | 1.29 | 0.912 | 2.35 |
| Dieldrin | 21.5 | 21.5 | 21.5 | 21.5 |
| Endrin | 20.7 | 20.8 | 20.7 | 20.8 |
| Endrin ketone | 8.5 | 8.5 | 8.5 | 8.5 |
| Sum DDD | 2,460 | 2,460 | 114 | 114 |
| Sum DDE | 906 | 906 | 906 | 906 |
| Sum DDT | 8,110 | 8,110 | 8,110 | 8,110 |
| Total chlordane | 8.1 | 8.1 | 669 | 669 |
| Total DDx | 11,500 | 11,500 | 11,500 | 11,500 |
| **Conventionals** | | | | |
| Ammonia (mg/kg) | 171 | 276 | 171 | 171 |
| Sulfide (mg/kg) | 29.1 | 38.5 | 29.1 | 38.5 |
| Total organic carbon (%) | 80.4 | 100 | 80.4 | 100 |
| Total % fines | 2.7 | 13 | 2.7 | 13 |

DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl

DDT – dichlorodiphenyltrichloroethane
dw – dry weight
EPA – US Environmental Protection Agency
FPM – floating percentile model
HPAH – high-molecular-weight polycyclic aromatic hydrocarbon
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
SQG – sediment quality guideline
SVOC – semivolatile organic compound
Total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT, and 4,4′-DDT)

Although an SQG was developed for each chemical provided to the FPM for each pooled endpoint, not all chemicals were needed to achieve acceptable error and reliability rates for any of the models. For the low toxicity thresholds, only six (Calcasieu), seven (EPA 2009), or nine (draft BERA) chemicals were needed. For the high toxicity thresholds, only ten (Calcasieu), six (EPA 2009), or four (draft BERA) were needed.

Chemicals that could be removed from a set without creating any change in the SQGs of other chemicals or the acceptability rates for that set could usually be removed because either: 1) the SQG for the chemical was equal to the maximum concentration observed in the Study Area for that chemical, indicating that there was no relationship between the chemical and toxicity in this dataset; or 2) the SQG was equal to the maximum no-hit concentration (i.e., the apparent effects threshold [AET]) for that chemical, and the chemical was redundant or correlated with other chemicals such that the false negative rate was not affected by the removal of the chemical.

Although only a small number of chemicals was needed to explain toxicity for each pooled endpoint with acceptable model error and reliability rates, there was little overlap among models with regard to the chemicals required. In order to create comparability among SQG sets, chemicals that were needed for the prediction of any pooled endpoint were retained for all models. Some model-specific exceptions were made in cases where the inclusion or exclusion of a chemical caused the model to fail performance criteria for false positive rates, false negative rates, and overall reliability. The rationale for retaining or eliminating individual chemicals from the final list of SQGs is provided in Table 4-4. Rationales for exclusion are marked with an “X,” and reasons for retaining the chemicals are briefly explained.

.

**Table 4-4. Chemicals Removed from Initial Sets of SQG Chemicals and Rationales for Removal**

| | Calcasieu | | | | | | EPA 2009 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Low SQGs | | | High SQGs | | | Low SQGs | | | High SQGs | | |
| **Chemical** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** |
| **Metals (mg/kg dw)** | | | | | | | | | | | | |
| Antimony | X | X | X | X | X | X | | X | | X | X | X |
| Arsenic | X | X | X | X | X | X | X | X | X | X | X | X |
| Cadmium[a] | Retained based on Calcasieu model | | | No effect – retained for comparison with other models | | | Retained based on EPA 2009 model | | | No effect – retained for comparison with other models | | |
| Chromium | X | X | X | X | X | X | X | X | X | X | X | X |
| Copper[a] | No effect – retained for comparison with draft BERA low toxicity results | | | No effect – retained for comparison with draft BERA low toxicity results | | | No effect – retained for comparison with draft BERA low toxicity results | | | No effect – retained for comparison with draft BERA low toxicity results | | |
| Lead | X | X | X | Retained only for Calcasieu model[b] | | X | X | X | X | X | X | X |
| Mercury[a] | No effect – retained for comparison with other models | | | Retained based on Calcasieu model | | | Retained based on EPA 2009 model | | | No effect – retained for comparison with other models | | |
| Silver | No effect – retained for comparison with EPA 2009 low toxicity results | | | No effect – retained for comparison with EPA 2009 low toxicity results | | | Retained based on EPA 2009 model | | | No effect – retained for comparison with EPA 2009 low toxicity results | | |
| Zinc | X | X | X | Retained only for Calcasieu model | | X | X | X | X | X | X | X |
| **Butyltins** | | | | | | | | | | | | |
| Dibutyltin ion | X | X | X | X | X | X | X | X | X | X | X | X |
| **PAHs (µg/kg)** | | | | | | | | | | | | |
| Total benzofluoranthenes[b] | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | |
| Total HPAHs | No effect – retained for comparison with draft BERA low toxicity results | | | No effect – retained for comparison with draft BERA low toxicity results | | | Retained based on EPA 2009 model | | | No effect – retained for comparison with draft BERA low toxicity results | | |

**Table 4-4. Chemicals Removed from Initial Sets of SQG Chemicals and Rationales for Removal**

| | Calcasieu | | | | | | EPA 2009 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Low SQGs | | | High SQGs | | | Low SQGs | | | High SQGs | | |
| **Chemical** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** |
| Total LPAHs[a, c] | Retained based on Calcasieu model | | | Retained based on Calcasieu model | | | Retained based on EPA 2009 model | | | Retained based on EPA 2009 model | | |
| Total PAHs | Removed – redundant with LPAH and HPAH | | | Removed – redundant with LPAH and HPAH | | | Removed – redundant with LPAH and HPAH | | | Removed – redundant with LPAH and HPAH | | |
| **SVOCs (µg/kg)** | | | | | | | | | | | | |
| Benzyl alcohol[b] | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | |
| Carbazole | No effect – retained for comparison with Calcasieu high toxicity model | | | Retained based on Calcasieu model | | | No effect – retained for comparison with Calcasieu high toxicity model | | | No effect – retained for comparison with Calcasieu high toxicity model | | |
| **Phenols (µg/kg)** | | | | | | | | | | | | |
| 4-Methyl phenol[d] | X | | X | X | | X | X | | X | Retained only for this model – needed for model performance | | X |
| Phenol | No effect – retained for comparison with draft BERA low toxicity results | | | No effect – retained for comparison with draft BERA low toxicity results | | | Retained based on EPA 2009 model | | | No effect – retained for comparison with draft BERA low toxicity results | | |
| **Phthalate (µg/kg)** | | | | | | | | | | | | |
| Diethyl phthalate[b] | X | | X | X | X | X | X | | | X | X | X |
| **PCBs** | | | | | | | | | | | | |
| Total PCBs (µg/kg)[a] | Retained based on Calcasieu model | | | No effect – retained for comparison with other models | | | Retained based on EPA 2009 model | | | Retained based on EPA 2009 model | | |
| **Pesticides (µg/kg)** | | | | | | | | | | | | |
| delta-Hexachlorocyclohexane[a] | Retained based on Calcasieu model | | | Retained based on Calcasieu model | | | Retained based on EPA 2009 model | | | No effect – retained for comparison with other models | | |

**Table 4-4. Chemicals Removed from Initial Sets of SQG Chemicals and Rationales for Removal**

| | Calcasieu | | | | | | EPA 2009 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Low SQGs | | | High SQGs | | | Low SQGs | | | High SQGs | | |
| **Chemical** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** | **Removal No Effect** | **SQG > Max** | **AET = Max** |
| Dieldrin[b] | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | |
| Endrin | No effect – retained for comparison with other models | | | Retained based on Calcasieu model | | | No effect – retained for comparison with other models | | | No effect – retained for comparison with other models | | |
| Endrin ketone[b] | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | | No effect – retained for comparison with draft BERA results | | |
| Sum DDD | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | |
| Sum DDE | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | |
| Sum DDT | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | | Removed – redundant with total DDx | | |
| Total Chlordane | X | X | X | X | | X | X | X | X | X | | X |
| Total DDx[a, c] | Retained based on Calcasieu model | | | Retained based on Calcasieu model | | | Retained based on EPA 2009 model | | | Retained based on EPA 2009 model | | |
| **Conventionals** | | | | | | | | | | | | |
| Ammonia (mg/kg)[a] | Retained based on Calcasieu model | | | No effect – retained for comparison with other models | | | Retained based on EPA 2009 model | | | Retained based on EPA 2009 model | | |
| Sulfide (mg/kg)[a, c] | Retained based on Calcasieu model | | | Retained based on Calcasieu model | | | No effect – retained for comparison with other models | | | Retained based on EPA 2009 model | | |
| Total organic carbon (%) | Retained based on Calcasieu model | | | Retained based on Calcasieu model | | | No effect – retained for comparison with other models | | | No effect – retained for comparison with other models | | |
| Total % fines | Retained only for this model – could not be removed | | X | X | X | X | Retained only for this model – could not be removed | | X | X | X | X |

a Required in draft BERA low toxicity endpoint.

[b] Included though not needed in the model.
[c] Required in BERA high toxicity endpoint.
[d] It was not possible to remove 4-methyl phenol from this model without retaining lead and zinc.

AET – apparent effects threshold
BERA – baseline ecological risk assessment
DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
DDT – dichlorodiphenyltrichloroethane
EPA – US Environmental Protection Agency
HPAH – high-molecular-weight polycyclic aromatic hydrocarbon
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl
SQG – sediment quality guideline
SVOC – semivolatile organic compound
total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT, and 4,4′-DDT)

The final sets of site-specific SQGs based on the pooled endpoints and the Calcasieu and EPA 2009 approaches are presented in Table 4-5. For three chemicals (mercury – Calcasieu approach, carbazole – Calcasieu approach, and sulfides – EPA 2009 approach) the low SQGs were greater than the high SQGs; therefore, the low SQGs were adjusted to be the same value as the high SQG. Final reliability rates are presented in Table 4-6.

**Table 4-5. Final Sets of High and Low SQGs**

| | Calcasieu SQGs | | BERA SQGs | | EPA 2009 SQGs | |
|---|---|---|---|---|---|---|
| **Chemical** | **Low Toxicity Threshold** | **High Toxicity Threshold** | **Low Toxicity Threshold** | **High Toxicity Threshold** | **Low Toxicity Threshold** | **High Toxicity Threshold** |
| **Metals (mg/kg dw)** | | | | | | |
| Cadmium | 0.714 | 3.51 | 0.714 | 3.51 | 0.507 | 3.51 |
| Copper | 562 | 562 | 76.9 | 562 | 493 | 562 |
| Mercury | 0.407 | 0.407 | 0.155 | 0.722 | 0.155 | 0.407 |
| Silver | 1.72 | 1.72 | 1.72 | 1.72 | 0.277 | 1.72 |
| Lead | NA | 179 | NA | NA | NA | NA |
| Zinc | NA | 469 | NA | NA | NA | NA |
| **PAHs (µg/kg)** | | | | | | |
| Total benzofluoranthenes | 53,000 | 53,000 | 53,000 | 53,000 | 53,000 | 53,000 |
| Total HPAHs | 610,000 | 610,000 | 610,000 | 610,000 | 22,000 | 610,000 |
| Total LPAHs | 2,300 | 33,000 | 2,000 | 18,000 | 1,600 | 9,300 |
| **SVOCs (µg/kg)** | | | | | | |
| 4-methyl phenol | NA | NA | NA | NA | NA | 96 |
| Benzyl alcohol | 36 | 36 | 36 | 36 | 36 | 36 |
| Carbazole | 540 | 540 | 1,100 | 1,100 | 1,100 | 1,100 |
| **Phthalates (µg/kg)** | | | | | | |
| Diethyl phthalate | 370 | NA | NA | NA | 120 | NA |
| **Phenols (µg/kg)** | | | | | | |
| Phenol | 120 | 120 | 120 | 120 | 34 | 120 |
| **PCBs (µg/kg)** | | | | | | |
| Total PCBs[a] | 300 | 3,500 | 300 | 500 | 250 | 500 |
| **Pesticides (µg/kg)** | | | | | | |
| delta-HCH | 1.29 | 1.29 | 1.29 | 2.35 | 1.26 | 2.35 |
| Dieldrin | 21.5 | 21.5 | 21.5 | 21.5 | 21.5 | 21.5 |
| Endrin | 20.7 | 20.8 | 20.7 | 20.8 | 20.7 | 20.8 |
| Endrin ketone | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 |
| Total DDx | 234 | 234 | 218 | 218 | 218 | 218 |
| **Conventionals** | | | | | | |
| Ammonia | 171 | 276 | 132 | 276 | 117 | 171 |
| Sulfide | 29.1 | 38.5 | 29.1 | 38.5 | 38.5 | 38.5 |
| Total organic carbon | 2.7 | NA | NA | NA | 13 | NA |

Sources: MacDonald Environmental (2002), Windward (2009), EPA (2009a, b)

[a] Total PCBs are based on Aroclors.

dw – dry weight
EPA – US Environmental Protection Agency
HCH – hexachlorocyclohexane
HPAH – high-molecular-weight polycyclic aromatic hydrocarbon
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
NA – not available
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl
SQG – sediment quality guideline
SVOC – semivolatile organic compound
Total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT, and 4,4′-DDT)

**Table 4-6. Final Reliability Rates Based on Pooled Endpoint Using SQGs in Table 4-5**

| | Low Risk Thresholds | | High Risk Thresholds | |
|---|---|---|---|---|
| **Reliability Parameter** | **Calcasieu** | **EPA 2009** | **Calcasieu** | **EPA 2009** |
| % False negatives | 18.64 | 19.44 | 18.42 | 18.18 |
| % False positives | 33.76 | 48.65 | 11.37 | 18.88 |
| % Sensitivity | 81.36 | 80.56 | 81.58 | 81.82 |
| % Efficiency | 66.24 | 51.35 | 88.63 | 81.12 |
| % Predicted hit reliability | 37.80 | 49.15 | 51.67 | 43.37 |
| % Predicted no-hit reliability | 93.37 | 81.90 | 97.00 | 96.19 |
| % Overall reliability | 69.28 | 62.12 | 87.71 | 81.23 |

EPA – US Environmental Protection Agency
SQG – sediment quality guideline

# 5.0 GENERIC SEDIMENT QUALITY GUIDELINES

The different methods of classifying samples as toxic or non-toxic also affected the performance of generic SQGs evaluated in the draft BERA (Windward 2009) for predicting site-specific toxicity. A brief summary of the effects is provided in this section.

## 5.1 EVALUATION OF GENERIC SQGS

Table 5-1 identifies the generic SQGs that were evaluated in the draft BERA (Windward 2009) and provides a brief description.

**Table 5-1. Description of Generic SQGs Evaluated in the Draft BERA**

| SQG | Toxicity Threshold | Description | Source |
|---|---|---|---|
| TEC | Low | The consensus-based TECs were intended to identify the concentration of sediment-associated contaminants below which adverse effects on sediment-dwelling organisms are not expected to occur. The TECs were derived as geometric means of SQGs (including TELs and ERLs) in the literature with similar narrative intent. The SQGs were derived from a combination of freshwater and marine toxicity tests. | Macdonald et al. (2000) |
| TEL | Low | The TELs were intended to estimate the concentrations of chemicals below which adverse biological effects only rarely occurred. The TELs were derived by calculating the geometric mean of the 15$^{th}$ percentile of the effect dataset and the 50$^{th}$ percentile of the no-effect dataset. These SQGs were derived from a national biological effects database that includes freshwater toxicity tests and changes in freshwater benthic community structure. | Smith et al. (1996) |
| SL1 | Low | The SL1s correspond to concentrations below which adverse effects on benthic organisms would not be expected. The SL1s were derived using the FPM and freshwater toxicity tests with both mortality and biomass endpoints. These SQGs were derived based on a regional (Western Washington and Oregon) biological effects database. | RSET (2006) |
| ERL | Low | The concentrations below the ERLs represent minimal-effects thresholds that were intended to estimate conditions below which effects would be rarely observed. The ERLs were derived as the 10$^{th}$ percentile of a database composed of multiple studies, species, and effects endpoints (predominantly mortality). These SQGs were derived from a national biological effects database that includes other sediment SQGs, marine toxicity tests, and benthic field studies. | Long et al. (1995) |
| SQS | Low | The SQS correspond to a concentration that will result in no adverse effects, including no acute or chronic adverse effects on biological resources and no significant health risk to humans. SQS are generally based on the LAETs (see LAET). These SQGs were derived from a biological effects database from Puget Sound, Washington (for estuarine, not freshwater sediments). | Ecology (1995) |

**Table 5-1. Description of Generic SQGs Evaluated in the Draft BERA**

| SQG | Toxicity Threshold | Description | Source |
|---|---|---|---|
| LAET | Low | The AET is the sediment contaminant concentration above which statistically significant ($p \leq 0.05$) adverse biological effects (relative to appropriate reference conditions) would always be expected. The LAET is the lowest of four AETs derived from four marine toxicity tests (amphipod mortality, echinoderm and oyster abnormality, and bacterial luminescence) and benthic community data. The biological effects database is from Puget Sound, Washington. | Gries and Waldow (1996) |
| PEC | High | The consensus-based PECs were intended to define the concentration of sediment-associated contaminants above which adverse effects on sediment-dwelling organisms are likely to occur. The PECs were derived as geometric means of SQGs (including ERMs and PELs) in the literature with similar narrative intent. The SQGs were derived from a combination of freshwater and marine toxicity tests. | MacDonald et al. (2000) |
| PEL | High | The PELs were intended to estimate the concentration of a chemical above which adverse biological effects frequently occurred. The PELs were derived by calculating the geometric mean of the 50$^{th}$ percentile of the effect dataset and the 85$^{th}$ percentile of the no-effect dataset. These SQGs were derived from a national biological effects database that includes freshwater toxicity tests and changes in freshwater benthic community structure. | Smith et al. (1996) |
| SL2 | High | The SL2 corresponds to a concentration at which minor adverse effects may be observed in the more sensitive groups of benthic organisms. SL2s were derived using the FPM and freshwater toxicity tests with both mortality and biomass endpoints. These SQGs were derived from a regional (Western Washington and Oregon) biological effects database. | RSET (2006) |
| ERM | High | The concentrations equivalent to or above the ERMs represent thresholds above which effects would frequently occur. The ERMs were derived as the 50$^{th}$ percentile of a database composed of multiple studies, species, and effects endpoints (predominantly mortality). These SQGs were derived from a national biological effects database that includes other sediment SQGs, marine toxicity tests, and benthic field studies. | Long et al. (1995) |
| CSL | High | The CSLs establish a minor adverse effects threshold above which adverse effects are expected to occur. CSLs are generally based on the second lowest AETs (see LAET). CSLs are used to define potential cleanup areas to be remediated under SMS. These SQGs are derived from a biological effects database from Puget Sound, Washington (for estuarine, not freshwater sediments). | Ecology (1995) |

Sources: MacDonald Environmental (2002), Windward (2009), EPA (2009a, b)

AET – apparent effects threshold
BERA – baseline ecological risk assessment
CSL – cleanup screening level
ERL – effects range – low
ERM – effects range – median
RSET – Regional Sediment Evaluation Team
SL1 – screening level 1
SL2 – screening level 2
SMS – Washington State Sediment Management Standards
SQG – sediment quality guideline

FPM – floating percentile model
LAET – lowest apparent effects threshold
PEC – probable effects concentration
PEL – probable effects level
SQS – sediment quality standards
TEC – threshold effects concentration
TEL – threshold effects level

The chemical screening process in the draft BERA (Windward 2009) identified 70 chemicals of potential concern (COPCs) for benthic invertebrates. The generic SQGs, mean quotients, and equilibrium partitioning sediment benchmarks (ESBs) for polycyclic aromatic hydrocarbon (PAH) mixtures, neutral organic compounds, and pesticides for the 70 COPCs were evaluated to identify chemicals or locations in the Study Area that might cause adverse effects on benthic organisms. As with the site-specific SQGs, the reliability of each set of generic SQGs, mean quotients, and ESBs was assessed for the two sets of toxicity thresholds (Calcasieu and EPA 2009) using the same reliability criteria of < 20% false negative and false positive rates, > 80% overall reliability, and > 90% predicted no hit reliability.

As was the case with the draft BERA hit thresholds (Windward 2009), none of the five sets of high SQGs met the reliability criteria when the Calcasieu or EPA 2009 hit thresholds were used, so the generic high SQGs were not used to predict benthic toxicity from surface sediment chemistry data. Similarly, the high SQG mean quotients and the ESBs for PAH mixtures, the sum of PAHs and narcotic non-ionic organics, and pesticides could not meet the acceptability criteria for the two high toxicity thresholds. Appendix D presents the reliability assessment that resulted in the rejection of these sets of generic SQGs for predicting benthic toxicity in the Study Area. Three sets of low SQGs (TECs, TELs, and ERLs) met the acceptability criterion for false negative rates, but not the criterion for false positive rates, for both the Calcasieu and EPA 2009 thresholds. Given that the intent of the low generic SQGs is to define non-toxic areas, the false positive error rate criterion was suspended and the low generic SQGs were applied to Study Area surface sediment chemistry data to identify non-toxic areas, as in was done in the draft BERA.

## 5.2 RISK CHARACTERIZATION USING LOW GENERIC SQGS

Three sets of low SQGs (TECs, TELs, and ERLs) met the reliability criteria for each of three approaches (Calcasieu, draft BERA, and EPA 2009). Maps 5-1 through 5-3 present the sediment chemistry stations in the Study Area identified as having no adverse effects on the benthic community based on the non-exceedance of one or more sets of these low SQGs. The results at these stations provide a conservatively biased estimate (i.e., statistically biased to underestimate the area) of non-toxic sediments.

# 6.0 COMPARISON OF APPROACHES TO ASSESSING BENTHIC RISK

Each SQG set was used to create a map displaying exceedances areas, as was done in the draft BERA (Windward 2009). Sediment concentrations were interpolated across a grid using a natural neighbors approach. This method does not distinguish the detection status of the empirical data used to interpolate intermediated sediment concentrations within a grid and thus may overestimate concentrations if a particular chemical has elevated detection limits (i.e., greater than the maximum detected value).

Maps 6-1 through 6-3 present the site-specific SQG exceedance areas based on the SQGs derived by the Calcasieu, draft BERA, and the EPA 2009 approaches. In general, the locations of the exceedance areas are similar among approaches, although the surface area associated with each exceedance area varies by method. Typically, the SQG exceedances derived from the Calcasieu approach define the smallest area, the SQG exceedance areas based on the EPA 2009 approach delineate the largest area, the area predicted to exceed site-specific SQGs in the draft BERA falls in between. The total area associated with SQG exceedances ranges from 121 acres (Calcasieu approach) to 193 acres (EPA 2009 approach). The estimate in the draft BERA was 165 acres. Based on this, the LWG recommends that the conclusions of the draft BERA be used to define AOPCs in the Feasibility Study, because they represent the middle of the range of uncertainty about benthic toxicity and benthic community risk.

Polar organic chemicals (benzyl alcohol, 4-methyl phenol, and phenol) were not used in the spatial analysis of SQG exceedances. These chemicals generally had low frequencies of detection and high detection limits and did not provide a reasonable basis for spatially representing benthic community effects. Predictions of toxicity based on high detection limits tended to identify unique areas that were not supported by empirical toxicity data. Ammonia and sulfides in sediment were also only measured at a subset of sediment sampling locations and could not be accurately represented spatially; thus these two conventional parameters were not included in the spatial analysis.

Several of the areas predicted to exceed site-specific SQGs under all approaches were small and therefore are unlikely to represent a risk to the benthic community. The Calcasieu approach delineates the greatest number of small toxic areas that were not identified by other methods because the Calcasieu method yielded SQGs for several metals (lead and zinc) not identified in other approaches. In the draft BERA (Windward 2009), these small areas were considered unlikely to represent risk to the benthic community. Remaining areas were considered potential benthic risk areas (PBRAs) and in the draft BERA were evaluated in terms of concordance between empirical toxicity data and predicted toxicity in individual PBRAs to arrive at final benthic community risk conclusions. The evaluation encompassed several steps; outcomes were used to address uncertainties regarding predicted unacceptable benthic risks within a PBRA (Map 6-4):

- The paired empirical toxicity test and chemistry results associated with each PBRA were identified. The data were sorted by toxicity status: toxic (i.e., Level 3 exceedance) versus non-toxic (Levels 0 or 1).

- For each chemical that exceeded its respective SQG for a given PBRA, characteristic chemical values of the toxic and non-toxic empirical sample subsets were identified:

    - The maximum concentration of the non-toxic samples

    - The lowest concentration that exceeded the SQG at a toxic station, where all toxic samples exceeded the SQG

    - If toxic sample concentrations were distributed both above and below the SQG, the SQG was retained as the best estimate of the minimum toxic sample concentration for a given chemical.

- If all bioassays within the PBRA were toxic, then the prediction of probable benthic risk was retained.

- If all bioassays within a PBRA were non-toxic, then the maximum concentration associated with the non-toxic samples was used to identify subareas that were unlikely to represent benthic risks.[9] If the maximum concentration within the PBRA occurred at a non-toxic station, the entire PBRA was identified as unlikely to represent an unacceptable benthic risk.

- If all toxic samples exceeded the SQG and the toxic and non-toxic sample chemistry overlapped (i.e., the minimum concentration in the toxic samples was less than the maximum concentration associated with the non-toxic samples), then both the minimum toxic and maximum non-toxic sample concentrations were used to refine risks within the PBRA.

    - An area with chemical concentrations below the minimum toxic sample concentration was considered unlikely to represent an unacceptable risk to the benthic community.

    - An area with the interpolated chemistry greater than the maximum non-toxic sample concentration continued to be represent unacceptable benthic risk.

[9] To identify subareas that may not represent a risk, the maximum concentration was re-applied to the natural neighbors-interpolated data within the specific PBRA and recontoured. Sediment chemical concentrations that exceeded the maximum non-toxic (≤ low reference thresholds defined by the reference envelope approach) sample concentration were considered to represent a risk.

  - An area with chemistry between the minimum toxic sample concentration (or SQG) and the maximum non-toxic sample concentration was considered uncertain.

- PBRAs that did not contain empirical bioassays were retained but were considered uncertain.

When this analysis was repeated based on SQGs derived using the Calcasieu and EPA 2009 REVs and effect thresholds, all (61) of the PBRAs identified in the BERA were also identified by the Calcasieu and EPA 2009 approaches, albeit with slightly different areas associated with individual PBRAs. Several new PBRAs were identified but represented very small, isolated areas that don't pose risks to the benthic community. Many of these new areas did not contain bioassay stations that could be used to corroborate SQG exceedances.

Overall, the three reference envelope approaches illustrate the uncertainty in predicting toxicity to benthic organisms exposed to Study Area sediments. Unfortunately, the *Hyalella* biomass hit classifications using the EPA 2009 approach were sufficiently uncertain – due to experimental errors unrelated to sediment toxicity – that toxicity could not reliably be predicted from sediment chemistry, with the consequence that the risk characterization has to rely on the other three bioassay endpoints when the EPA 2009 hit thresholds are used.

The degree of consistency in the results across endpoints and methods is very encouraging. The SQG exceedance areas varied somewhat by method of deriving hit thresholds (from 121 acres by the Calcasieu method, to 165 acres by the draft BERA method, to 193 acres by the EPA 2009 method), but the three methods are remarkably consistent and generally identify the same exceedance areas. The level of consistency noted across the three methods should provide confidence that robust risk management decisions may be made despite uncertainty regarding precise bioassay hit thresholds.

# 7.0 REFERENCES

Anderson MR. 2008. Description and use of the RSET floating percentile method spreadsheets. Draft. September 6, 2008. Regional Sediment Evaluation Team, Oregon Department of Environmental Quality, Portland, OR.

ASTM. 2007. Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. Method E1706-00. 2007 Annual book of ASTM standards, vol 11.06 [online]. American Society for Testing and Materials, West Conshohocken, PA. Available from: http://www.astm.org/cgi-bin/SoftCart.exe/DATABASE.CART/REDLINE_PAGES/E1706.htm?L+mystore+xzea3685+1187379245.

Avocet. 2003. Development of freshwater sediment quality values for use in Washington State. Phase II report: Development and recommendation of SQVs for freshwater sediments in Washington State. Publication No. 03-09-088. Prepared for Washington Department of Ecology. Avocet Consulting, Kenmore, WA.

Ecology. 1995. Sediment management standards. WAC 173-204. Washington State Department of Ecology, Toxics Cleanup Program, Sediment Management Unit, Olympia, WA.

EPA. 2006. EPA comments on ecological risk assessment interpretive report: estimating risks to benthic organisms using predictive models based on sediment toxicity tests, and accompanying cover letter to Lower Willamette Group from C. Humphrey and E. Blischke dated July 6, 2006. US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2008. EPA e-mail dated October 14, 2008 to Lower Willamette Group (from E. Blischke to R. Wyatt) regarding benthic evaluation for Portland Harbor. US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2009a. EPA e-mail dated July 17, 2009, to Lower Willamette Group (from E. Blischke to R. Wyatt) regarding Portland Harbor RI/FS: interpretation of sediment bioassay results. US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2009b. EPA e-mail dated July 31, 2009, to Lower Willamette Group (from E. Blischke to J. McKenna) regarding Portland Harbor RI/FS: benthic interpretation. US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2009c. EPA e-mail dated March 31, 2009 to Lower Willamette Group (from E. Blischke to J. Toll) regarding development of reference envelope for the evaluation of benthic risk, with two attached files. US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

EPA. 2009d. EPA letter dated August 19, 2009 to Lower Willamette Group (from E. Blischke and C. Humphrey to R. Wyatt) regarding Portland Harbor Superfund site: submittal of baseline ecological risk assessment. US Environmental Protection Agency Region 10, Oregon Operations Office, Portland, OR.

Gries TH, Waldow KH. 1996. Progress re-evaluating Puget Sound apparent effects thresholds (AETs). Volume I: 1994 amphipod and echinoderm larval AETs. Draft report. Prepared for Puget Sound Dredged Disposal Analysis (PSDDA) agencies. Washington Department of Ecology, Olympia, WA.

Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. Environ Manage 19:81-97.

MacDonald DD, Landrum PF. 2008. An evaluation of the approach for assessing risks to the benthic invertebrate community at the Portland Harbor Superfund site. Preliminary draft. Prepared for US Environmental Protection Agency. September 2008. MacDonald Environmental Sciences, Ltd., Nanaimo, BC, and Landrum and Associates, Ann Arbor, MI.

MacDonald DD, Ingersoll CG, Berger TA. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39(5):20-31.

MacDonald Environmental. 2002. Calcasieu Estuary remedial investigation/feasibility study (RI/FS): Baseline ecological risk assessment (BERA). Prepared for CDM Federal Programs Corp. September 2002. MacDonald Environmental Services, Ltd., Nanaimo, BC.

RSET. 2006. Regional sediment evaluation framework for the Pacific Northwest. Interim final. September 2006. Regional Sediment Evaluation Team (US Army Corps of Engineers - Seattle District, Portland District, Walla Walla District, and Northwestern Division; US Environmental Protection Agency Region 10; Washington Department of Ecology; Washington Department of Natural Resources; Oregon Department of Environmental Quality; Idaho Department of Environmental Quality; National Marine Fisheries Service; US Fish and Wildlife Service).

RSET. 2009. Regional sediment evaluation framework for the Pacific Northwest. Final. May 2009. Regional Sediment Evaluation Team (US Army Corps of Engineers - Seattle District, Portland District, Walla Walla District, and Northwestern Division; US Environmental Protection Agency Region 10; Washington Department of Ecology; Washington Department of Natural Resources; Oregon Department of Environmental Quality; Idaho Department of Environmental Quality; National Marine Fisheries Service; US Fish and Wildlife Service).

Smith SL, MacDonald DD, Keenleyside KA, Ingersoll CG, Field LJ. 1996. A preliminary evaluation of sediment quality assessment values for freshwater ecosystems. J Great Lakes Res 22:624-638.

Toll J. 2009a. E-mail dated August 26, 2009 to EPA (B. Shephard, C. Humphrey, E. Blischke) regarding record of decisions from 8-26-09 meeting to define EPA-approved bioassay reference envelope values, with attached Excel file EPA bioassay hit thresholds 082609.xlsx. Windward Environmental, Seattle, WA.

Toll J. 2009b. E-mail dated August 27, 2009, to LWG (J. Woronets, B. Wyatt, J. McKenna, K. Pine,) regarding response of Burt Shephard to 8/26/09 e-mail regarding record of decisions from 8/26/09 meeting. Windward Environmental, Seattle, WA.

Windward. 2009. Portland Harbor RI, Appendix G: Baseline ecological risk assessment. Draft. Prepared for the Lower Willamette Group. August 19, 2009. Windward Environmental LLC, Seattle, WA.

Windward, Avocet, TerraStat. 2005. Memorandum to Lower Willamette Group: Benthic interpretive approach for Portland Harbor Ecological Risk Assessment. December 28, 2005. Windward Environmental LLC, Seattle, WA; Avocet Consulting, Kenmore, WA; TerraStat Consulting Group, Snohomish, WA.

PORTLAND HARBOR RI

# BENTHIC TOXICITY REANALYSIS TECHNICAL MEMORANDUM

## APPENDIX A
## EPA AND CALCASIEU THRESHOLD CALCULATIONS

**DRAFT**

**DO NOT QUOTE OR CITE**
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

November 13, 2009

**Prepared for**
The Lower Willamette Group

**Prepared by**
Windward Environmental LLC

# TABLE OF CONTENTS

PORTLAND HARBOR RI

# BENTHIC TOXICITY REANALYSIS TECHNICAL MEMORANDUM

## APPENDIX B
## FLOATING PERCENTILE MODEL

**DRAFT**

| |
|---|
| **DO NOT QUOTE OR CITE**<br>This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part. |

November 13, 2009

**Prepared for**
The Lower Willamette Group

**Prepared by**
Windward Environmental LLC

# TABLE OF CONTENTS

# LIST OF TABLES

# 1.0 RELIABILITY ANALYSIS OF THE FPM

The selection of the sediment quality guidelines (SQGs) from the floating percentile model (FPM) involved a series of model runs and reliability analyses. The reliability analyses are coded in the current version of the FPM software (Anderson 2008), which was used for the draft baseline ecological risk assessment (BERA). For each risk level, the first step involved running the FPM for individual endpoints, selecting the minimum sediment quality guideline (SQG) for each chemical from those endpoints, and testing those minimum SQGs against the pooled endpoint. This appendix describes that process.

The FPM was run for each of the eight endpoints – *Chironomus* and *Hyalella* low- and high-risk survival and biomass – using two sets of toxicity thresholds, Calcasieu and US Environmental Protection Agency (EPA) 2009, for a single false negative rate of 20%. Tables 1-1 and 1-2 present the reliability results for each of the endpoints for the Calcasieu and EPA 2009 thresholds, respectively. The thresholds are defined as follows:

- **High toxicity thresholds:** Models developed for all endpoints, except the EPA 2009 *Hyalella* biomass endpoint, met the acceptability criteria. The *Hyalella* biomass high toxicity endpoint was excluded from further consideration in SQG development for the EPA 2009 set of endpoints.

- **Low toxicity thresholds:** Models developed for all endpoints using the Calcasieu thresholds passed all acceptability criteria. For the EPA 2009 thresholds, only *Hyalella* survival[1] endpoint achieved the desired criterion. Because the goal for these endpoints was to designate non-toxic areas, incorrectly predicting toxicity above the SQG was not considered as important as guaranteeing that false negative rates were maintained, so all endpoints were retained for SQG development.[2]

**Table 1-1. Reliability Results for Initial Endpoint-Specific FPM Runs Based on Calcasieu Thresholds**

| | Reliability Parameters | | | | | | |
|---|---|---|---|---|---|---|---|
| **Endpoint by Toxicity Threshold** | **% False Negatives** | **% False Positives** | **% Sensitivity** | **% Efficiency** | **% Predicted Hit Reliability** | **% Predicted No-Hit Reliability** | **% Overall Reliability** |
| **Low Toxicity Thresholds** | | | | | | | |
| *Chironomus* survival | 18.92 | 19.14 | 81.08 | 80.86 | 37.97 | 96.73 | 80.89 |
| *Chironomus* biomass | 18.18 | 15.26 | 81.82 | 84.74 | 48.65 | 96.35 | 84.30 |
| *Hyalella* survival | 19.05 | 4.41 | 80.95 | 95.59 | 58.62 | 98.48 | 94.54 |
| *Hyalella* biomass | 20.00 | 19.01 | 80.00 | 80.99 | 32.43 | 97.26 | 80.89 |
| **High Toxicity Thresholds** | | | | | | | |
| *Chironomus* survival | 17.86 | 11.32 | 82.14 | 88.68 | 43.40 | 97.92 | 88.05 |
| *Chironomus* biomass | 19.44 | 8.17 | 80.56 | 91.83 | 58.00 | 97.12 | 90.44 |

[1] Hits for Hyalella survival were identical using the Calcasieu and EPA thresholds.
[2] The high false positive rates for most low SQG sets had the effect of lowering the overall reliability rates below 80%.

**Table 1-1. Reliability Results for Initial Endpoint-Specific FPM Runs Based on Calcasieu Thresholds**

| | Reliability Parameters | | | | | | |
|---|---|---|---|---|---|---|---|
| **Endpoint by Toxicity Threshold** | **% False Negatives** | **% False Positives** | **% Sensitivity** | **% Efficiency** | **% Predicted Hit Reliability** | **% Predicted No-Hit Reliability** | **% Overall Reliability** |
| *Hyalella* survival | 16.67 | 3.64 | 83.33 | 96.36 | 60.00 | 98.88 | 95.56 |
| *Hyalella* biomass | 18.75 | 3.25 | 81.25 | 96.75 | 59.09 | 98.89 | 95.90 |

FPM – floating percentile model

**Table 1-2. Reliability Results for Initial Endpoint-Specific FPM Runs Based on EPA 2009 Thresholds**

| | Reliability Parameters | | | | | | |
|---|---|---|---|---|---|---|---|
| **Endpoint by Toxicity Threshold** | **% False Negatives** | **% False Positives** | **% Sensitivity** | **% Efficiency** | **% Predicted Hit Reliability** | **% Predicted No-Hit Reliability** | **% Overall Reliability** |
| **Low Toxicity Thresholds** | | | | | | | |
| *Chironomus* survival | 18.18 | 26.51 | 81.82 | 73.49 | 35.29 | 95.81 | 74.74 |
| *Chironomus* biomass | 19.23 | 21.16 | 80.77 | 78.84 | 45.16 | 95.00 | 79.18 |
| *Hyalella* survival | 19.05 | 4.41 | 80.95 | 95.59 | 58.62 | 98.48 | 94.54 |
| *Hyalella* biomass | 20.00 | 54.13 | 80.00 | 45.87 | 33.71 | 86.96 | 54.61 |
| **High Toxicity Thresholds** | | | | | | | |
| *Chironomus* survival | 18.75 | 12.26 | 81.25 | 87.74 | 44.83 | 97.45 | 87.03 |
| *Chironomus* biomass | 19.05 | 13.94 | 80.95 | 86.06 | 49.28 | 96.43 | 85.32 |
| *Hyalella* survival | 15.79 | 4.74 | 84.21 | 95.26 | 55.17 | 98.86 | 94.54 |
| *Hyalella* biomass | 17.95 | 33.86 | 82.05 | 66.14 | 27.12 | 96.00 | 68.26 |

EPA – US Environmental Protection Agency
FPM – floating percentile model

For each level of toxicity (i.e., low or high) and threshold (i.e., Calcasieu or EPA 2009), an initial set of SQGs was created by selecting the minimum SQG for each chemical from the SQGs for any individual endpoints that passed acceptability criteria at that level (Table 1-3). Acceptability criteria were then calculated using the minimum SQG sets to predict the pooled low and high endpoints for each threshold. None of the sets of minimum SQGs met the performance standards for predicting the pooled endpoints (Table 1-4), and so the FPM was run to develop a set of SQGs that would directly predict the pooled endpoints with the desired reliability and error rates.

**Table 1-3. Initial Set of FPM SQGs Based on Minimum SQG of Individual Endpoints that Passed Acceptability Criteria at Each Level**

| | Calcasieu | | EPA 2009 | |
|---|---|---|---|---|
| **Chemical** | **Low Toxicity Threshold** | **High Toxicity Threshold** | **Low Toxicity Threshold** | **High Toxicity Threshold** |
| **Metals (mg/kg dw)** | | | | |
| Antimony | 19.3 | 19.3 | 0.69 | 19.3 |
| Arsenic | 22.9 | 34 | 22.9 | 34 |
| Cadmium | 0.714 | 3.51 | 0.507 | 3.51 |
| Chromium | 224 | 81.6 | 224 | 81.6 |
| Copper | 562 | 562 | 493 | 562 |
| Lead | 179 | 179 | 1,290 | 179 |
| Mercury | 0.407 | 0.407 | 0.249 | 0.624 |
| Silver | 1.72 | 1.72 | 0.277 | 1.72 |
| Zinc | 1,940 | 1,940 | 1,940 | 1,940 |
| **Butyltins** | | | | |
| Dibutyltin ion | 150 | 910 | 380 | 910 |
| **PAHs (µg/kg)** | | | | |
| Total benzofluoranthenes | 53,000 | 53,000 | 53,000 | 53,000 |
| Total HPAHs | 610,000 | 610,000 | 22,000 | 610,000 |
| Total LPAHs | 2,300 | 18,000 | 1,600 | 2,300 |
| Total PAHs | 1,300,000 | 330,000 | 1,300,000 | 1,300,000 |
| **SVOCs (µg/kg)** | | | | |
| Benzyl alcohol | 36 | 36 | 16 | 36 |
| Carbazole | 1,100 | 540 | 1,100 | 540 |
| **Phenols (µg/kg)** | | | | |
| 4-Methyl phenol | 96 | 125 | 96 | 96 |
| Phenol | 120 | 120 | 32 | 120 |
| **Phthalate (µg/kg)** | | | | |
| Diethyl phthalate | 7.9 | 370 | 7.9 | 370 |
| **PCBs** | | | | |
| Total PCBs (µg/kg) | 3,500 | 3,500 | 500 | 3,500 |
| **Pesticides (µg/kg)** | | | | |
| delta-Hexachlorocyclohexane | 1.29 | 2.35 | 1.26 | 2.35 |
| Dieldrin | 21.5 | 21.5 | 21.5 | 21.5 |
| Endrin | 20.7 | 20.8 | 20.7 | 20.7 |
| Endrin ketone | 8.5 | 8.5 | 8.5 | 8.5 |
| Sum DDD | 114 | 331 | 114 | 331 |
| Sum DDE | 906 | 906 | 906 | 906 |
| Sum DDT | 8,110 | 8,110 | 32.5 | 71.6 |
| Total chlordane | 8.1 | 12 | 669 | 669 |
| Total DDx | 11,500 | 11,500 | 11,500 | 11,500 |

**Table 1-3. Initial Set of FPM SQGs Based on Minimum SQG of Individual Endpoints that Passed Acceptability Criteria at Each Level**

| | Calcasieu | | EPA 2009 | |
|---|---|---|---|---|
| **Chemical** | **Low Toxicity Threshold** | **High Toxicity Threshold** | **Low Toxicity Threshold** | **High Toxicity Threshold** |
| **Conventionals** | | | | |
| Ammonia (mg/kg) | 164 | 276 | 120 | 171 |
| Sulfide (mg/kg) | 38.5 | 38.5 | 38.5 | 38.5 |
| Total organic carbon (%) | 2.7 | 13 | 2.7 | 13 |
| Total % fines | 84.4 | 100 | 76.7 | 100 |

DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
DDT – dichlorodiphenyltrichloroethane
dw – dry weight
EPA – US Environmental Protection Agency
FPM – floating percentile model
HPAH – high-molecular-weight polycyclic aromatic hydrocarbon
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl
SQG – sediment quality guideline
SVOC – semivolatile organic compound
Total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT, and 4,4′-DDT)

**Table 1-4. Reliability Rates for Predicting the Pooled Endpoint Using Minimum SQGs from Individual Endpoints**

| | Low Risk Threshold | | High Risk Threshold | |
|---|---|---|---|---|
| **Reliability Parameter** | **Calcasieu** | **EPA 2009** | **Calcasieu** | **EPA 2009** |
| % False negatives | 22.03 | 18.52 | 21.05 | 41.54 |
| % False positives | 32.48 | 55.14 | 11.37 | 20.61 |
| % Sensitivity | 77.97 | 81.48 | 78.95 | 58.46 |
| % Efficiency | 67.52 | 44.86 | 88.63 | 79.39 |
| % Predicted hit reliability | 37.70 | 46.32 | 50.85 | 44.71 |
| % Predicted no-hit reliability | 92.40 | 80.58 | 96.58 | 87.02 |
| % Overall reliability | 69.62 | 58.36 | 87.37 | 74.74 |

EPA – US Environmental Protection Agency
SQG – sediment quality guideline

# 2.0 REFERENCES

Anderson MR. 2008. Description and use of the RSET floating percentile method spreadsheets. Draft. September 6, 2008. Regional Sediment Evaluation Team, Oregon Department of Environmental Quality, Portland, OR.

PORTLAND HARBOR RI

# BENTHIC TOXICITY REANALYSIS TECHNICAL MEMORANDUM

## APPENDIX C
## LOGISTIC REGRESSION MODEL

**DRAFT**

**DO NOT QUOTE OR CITE**
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

November 13, 2009

**Prepared for**
The Lower Willamette Group

**Prepared by**
Windward Environmental LLC

# TABLE OF CONTENTS

# LIST OF TABLES

# 1.0 THE LOGISTIC REGRESSION MODEL

The logistic regression model (LRM) estimates a functional curve relationship between concentrations of a chemical and the proportion of toxicity detected in the sediment toxicity tests (Field et al. 2002). This enables users to select the specific probability of effects based on a risk management decision that corresponds to their specific assessment objective.

A suite of individual LRMs were fit to each chemical after the exclusion of chemicals with fewer than 30 detected values and certain conventional parameters (e.g., specific gravity, liquid limit, individual grain size, ammonia, total organic carbon and total solids), which were not considered contaminants. Each individual model was reviewed, and all chemicals with a poorly fit model or insufficient data were omitted from the LRM development process. These were models that had Chi-square p-values greater than 0.01, $R^2_L$ values less than 0.20, or fewer than five toxic samples retained in the screened dataset. For each toxicity test endpoint, individual chemical models were scored based on predictive performance for the Portland Harbor draft baseline ecological risk assessment (BERA) dataset. Individual models with excessive false positives (false positives $\geq$ true positives) were excluded in an effort to limit erroneous predictions and reduce the error rate. The best suite of chemical models for each toxicity test endpoint was retained; this process did not restrict the final suite to the same list of chemicals for each toxicity test endpoint. Table 1-1 presents the chemicals retained in the LRM for one or more toxicity test endpoint.

**Table 1-1. Chemicals Included in the Logistic Regression Model**

| Chemical | Number of Detections | Chemical | Number of Detections |
|---|---|---|---|
| **Metals (mg/kg)** | | | |
| Antimony | 60 | Mercury | 206 |
| Arsenic | 56 | Selenium | 274 |
| Cadmium | 56 | Silver | 269 |
| Copper | 285 | Zinc | 58 |
| Lead | 196 | | |
| **PAHs (µg/kg)** | | | |
| 1-Methylnaphthalene | 293 | C3-Fluoranthene/pyrene | 60 |
| 2-Methylnaphthalene | 270 | C3-Fluorene | 36 |
| Acenaphthene | 59 | C4-Naphthalene | 57 |
| Acenaphthylene | 282 | Chrysene | 281 |
| Anthracene | 277 | Dibenzo(a,h)anthracene | 275 |
| Benzo(a)anthracene | 58 | Fluoranthene | 288 |
| Benzo(a)pyrene | 59 | Fluorene | 280 |
| Benzo(b)fluoranthene | 59 | Indeno(1,2,3-cd)pyrene | 58 |
| Benzo(g,h,i)perylene | 57 | Naphthalene | 206 |

**Table 1-1. Chemicals Included in the Logistic Regression Model**

| Chemical | Number of Detections | Chemical | Number of Detections |
|---|---|---|---|
| Benzo(k)fluoranthene | 60 | Perylene | 165 |
| C1-Dibenzothiophene | 60 | Phenanthrene | 231 |
| C1-Fluorene | 59 | Pyrene | 175 |
| C1-Naphthalene | 56 | Sum of 34 PAHs | 39 |
| C2-Fluorene | 56 | Total benzofluoranthenes | 192 |
| C2-Naphthalene | 58 | Total HPAHs | 202 |
| C2-Phenanthrene/anthracene | 57 | Total LPAHs | 59 |
| C3-Dibenzothiophene | 59 | Total PAHs | 34 |
| **SVOCs (µg/kg)** | | | |
| Benzyl alcohol | 59 | Dibenzofuran | 271 |
| Carbazole | 278 | | |
| **Phenol (µg/kg)** | | | |
| 4-Methyl phenol | 292 | Phenol | 108 |
| **Phthalates (µg/kg)** | | | |
| Bis(2-ethylhexyl) phthalate | 60 | Dibutyl phthalate | 232 |
| Butylbenzyl phthalate | 59 | | |
| **PCBs (µg/kg)** | | | |
| Aroclor 1254 | 7 | Total PCBs | 48 |
| Aroclor 1260 | 41 | Total PCB Aroclors | 31 |
| **PCB Congeners (pg/g)** | | | |
| PCB 001 | 293 | PCB 105 | 281 |
| PCB 002 | 70 | PCB 106.118 | 120 |
| PCB 003 | 74 | PCB 107.109 | 293 |
| PCB 011 | 34 | PCB 110 | 293 |
| PCB 025 | 73 | PCB 114 | 58 |
| PCB 026 | 293 | PCB 124 | 272 |
| PCB 035 | 224 | PCB 128.162 | 274 |
| PCB 037 | 293 | PCB 129 | 274 |
| PCB 081 | 291 | PCB 137 | 275 |
| PCB 084.092 | 292 | PCB 148 | 281 |
| PCB 087.117.125 | 293 | PCB 157 | 281 |
| PCB 097 | 293 | PCB 159 | 280 |
| PCB 099 | 289 | | |

**Table 1-1. Chemicals Included in the Logistic Regression Model**

| Chemical | Number of Detections | Chemical | Number of Detections |
|---|---|---|---|
| **Pesticides (µg/kg)** | | | |
| 2,4'-DDD | 293 | Dieldrin | 281 |
| 2,4'-DDE | 293 | Endrin | 282 |
| 2,4'-DDT | 293 | Endrin ketone | 285 |
| 4,4'-DDD | 292 | Sum DDD | 85 |
| 4,4'-DDE | 240 | Sum DDE | 56 |
| 4,4'-DDT | 293 | Sum DDT | 74 |
| cis-Chlordane | 57 | Total chlordane | 154 |
| alpha-Endosulfan | 60 | Total DDx | 92 |
| beta-HCH | 287 | Total endosulfan | 52 |
| delta-HCH | 283 | trans-Nonachlor | 70 |
| gamma-HCH | 288 | | |
| **Conventionals (mg/kg)** | | | |
| Sulfide | 223 | | |

DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
DDT – dichlorodiphenyltrichloroethane
HCH – hexachlorocyclohexane
HPAH – high-molecular-weight polycyclic aromatic hydrocarbon
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl
SVOC – semivolatile organic compound
Total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT and 4,4′-DDT)

Using the toxicity thresholds derived from the Calcasieu and EPA 2009 approaches, the LRM was used to derive site-specific SQGs. The selection of the SQGs from the LRM involved a series of model runs and subsequent reliability analyses. The reliability of the LRM was evaluated using the reliability goals for the set of SQGs presented in the draft *Development of Freshwater Sediment Quality Values for Use in Washington State* (Avocet 2003). Based on this document, both false negative and false positive rates should be below 20%, and the overall reliability should be above 80%. In addition, predicted no-hit reliability should be above 90% in order to have greater confidence in defining a sampling location as having no adverse effects.

In the *Portland Harbor RI/FS Comprehensive Round 2 Site Characterization Summary and Data Gaps Analysis Report* (Integral et al. 2007), total petroleum hydrocarbons (TPH) were identified as being correlated with the detected toxicity in the toxicity tests. The TPH data were later reclassified into pyrogenic (PYO) compounds, petrogenic (PTO) compounds, and other petroleum hydrocarbons; and subsequent analysis determined that PYO and PTO compounds were correlated with the detected toxicity. The same correlations were found in the draft BERA, Calcasieu, and EPA 2009 approaches for toxicity definitions. In an effort to separate out the predictive ability of the petroleum and non-petroleum compounds, two model sets were evaluated. The best-fit model for PYO and PTO compounds was retained as a separate model (Tables 1-2 and 1-3). All the other chemicals were combined into a multiple-chemical model, in which the maximum probability of toxicity ($max_p$) across all chemicals was calculated. Individual models with excessive false positives (false positives ≥ true positives) were excluded in an effort to limit erroneous predictions and reduce the error rate (Tables 1-4 and 1-5; this chemical list refinement was performed for the high toxicity thresholds only). As specified in EPA's problem formulation, as presented in the draft BERA (Windward 2009), the two probability (Pr) values of 0.4 and 0.6 were selected as the thresholds to predict non-toxic and toxic sediment samples, respectively.

None of the LRMs for the Calcasieu and EPA 2009 approaches met the reliability goals (Tables 1-2 through 1-5). All of the sampling locations that were predicted to be toxic based on TPH were also predicted to be toxic based on another chemical in the non-petroleum chemical set, so having separate predictions for TPH alone and for non-petroleum chemicals did not help to improve reliability. The TPH data were available for only 41 of the 293 stations, so a predictive model based on TPH compounds would have limited spatial relevance.

As noted above, the Pr thresholds for toxicity used in these LRM predictions were 0.4 and 0.6, as specified by EPA in the problem formulation as presented in the draft BERA (Windward 2009). Because the majority of the LRMs failed the criterion for the false negative rate, the Pr thresholds were lowered until this criterion was met. By making this adjustment, all LRMs failed to meet the other acceptability criteria; in addition, the Pr thresholds were extremely low (most Pr thresholds were ≤ 26%) such that they did not provide a reasonable separation between toxic and non-toxic samples.

The low Pr thresholds resulted from attempts to simultaneously keep both error rates below 20%. The dataset includes a number of low-concentration toxic sampling locations that drive the toxicity threshold down. Because there are so few toxic sampling locations, each sampling location contributes substantially to the error percentage and to the ultimate toxicity threshold needed to achieve the target error rates. Based on this reliability analysis, no site-specific SQGs suitable for predicting adverse effects to benthic invertebrates in the Lower Willamette River Study Area could be derived using the LRMs. Hence the LRM was not used to predict adverse effects to the benthic community in the benthic risk assessment.

**Table 1-2. LRM Reliability Results for the Calcasieu Toxicity Thresholds and Petrogenic and Pyrogenic Petroleum Compounds**

| | Reliability Parameters | | | | | | |
|---|---|---|---|---|---|---|---|
| **Endpoint by Toxicity Thresholds** | **% False Negatives** | **% False Positives** | **% Sensitivity** | **% Efficiency** | **% Predicted Hit Reliability** | **% Predicted No-Hit Reliability** | **% Overall Reliability** |
| **Low Toxicity Thresholds** | | | | | | | |
| *Chironomus* survival | 22 | 4 | 67 | 87 | 92 | 83 | 78 |
| *Chironomus* biomass | 32 | 11 | 64 | 79 | 88 | 68 | 71 |
| *Hyalella* survival | 53 | 8 | 40 | 92 | 75 | 75 | 73 |
| *Hyalella* biomass | 57 | 4 | 43 | 93 | 86 | 76 | 76 |
| **High Toxicity Thresholds** | | | | | | | |
| *Chironomus* survival | 36 | 4 | 64 | 93 | 90 | 83 | 83 |
| *Chironomus* biomass | 30 | 5 | 60 | 86 | 92 | 75 | 73 |
| *Hyalella* survival | 46 | 7 | 46 | 93 | 75 | 81 | 78 |
| *Hyalella* biomass | 45 | 3 | 55 | 93 | 86 | 85 | 83 |

LRM – logistic regression model

**Table 1-3. LRM Reliability Results for the EPA 2009 Toxicity Thresholds and Petrogenic and Pyrogenic Petroleum Compounds**

| | Reliability Parameters | | | | | | |
|---|---|---|---|---|---|---|---|
| **Endpoint by Toxicity Threshold** | **% False Negatives** | **% False Positives** | **% Sensitivity** | **% Efficiency** | **% Predicted Hit Reliability** | **% Predicted No-Hit Reliability** | **% Overall Reliability** |
| **Low Toxicity Thresholds** | | | | | | | |
| *Chironomus* survival | 26 | 5 | 63 | 86 | 92 | 79 | 76 |
| *Chironomus* biomass | 17 | 17 | 61 | 67 | 82 | 75 | 63 |
| *Hyalella* survival | 53 | 8 | 40 | 92 | 75 | 75 | 73 |
| *Hyalella* biomass | 70 | 5 | 30 | 90 | 86 | 58 | 61 |
| **High Toxicity Thresholds** | | | | | | | |
| *Chironomus* survival | 29 | 8 | 53 | 88 | 82 | 81 | 73 |
| *Chironomus* biomass | 33 | 10 | 62 | 80 | 87 | 70 | 71 |
| *Hyalella* survival | 50 | 7 | 43 | 93 | 75 | 78 | 76 |
| *Hyalella* biomass | 63 | 4 | 38 | 92 | 86 | 70 | 71 |

LRM – logistic regression model

**Table 1-4. LRM Reliability Results for the Calcasieu Toxicity Thresholds and Non-Petroleum Chemicals**

| Endpoint by Toxicity Thresholds | Number of Chemicals in LRM Set | Reliability Parameters | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | % False Negatives | % False Positives | % Sensitivity | % Efficiency | % Predicted Hit Reliability | % Predicted No-Hit Reliability | % Overall Reliability |
| **Low Toxicity Thresholds**[a] | | | | | | | | |
| *Chironomus* survival | 47 | 35 | 7 | 51 | 87 | 53 | 94 | 83 |
| *Chironomus* biomass | 59 | 25 | 13 | 64 | 78 | 47 | 95 | 76 |
| *Hyalella* survival | 40 | 33 | 4 | 57 | 90 | 55 | 97 | 88 |
| *Hyalella* biomass | 47 | 53 | 5 | 43 | 88 | 48 | 94 | 84 |
| **High Toxicity Thresholds**[b] | | | | | | | | |
| *Chironomus* survival | 45 | 32 | 5 | 54 | 89 | 52 | 96 | 85 |
| *Chironomus* biomass | 38 | 31 | 7 | 58 | 88 | 54 | 95 | 85 |
| *Hyalella* survival | 21 | 33 | 2 | 56 | 95 | 67 | 98 | 93 |
| *Hyalella* biomass[c] | 3 | 47 | 1 | 47 | 99 | 78 | 97 | 96 |

[a] Predictions are based on a reduced chemical list that excluded petroleum compounds, unreliable models, and conventionals.

[b] Predictions are based on a reduced chemical list that excluded petroleum compounds, unreliable models, conventionals, and chemicals with high false positive error rates within each toxicity endpoint.

[c] Only three chemicals were left in this LRM set: 2,4'-DDE, beta-HCH, and carbazole. As a result, toxicity predictions were not made for 21 stations because of missing values for all three of these chemicals.

HCH – hexachlorocyclohexane

LRM – logistic regression model

**Table 1-5. LRM Reliability Results for the EPA 2009 Toxicity Thresholds and Non-Petroleum Chemicals**

| Endpoint by Toxicity Threshold | Number of Chemicals in LRM Set | Reliability Parameters | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | % False Negatives | % False Positives | % Sensitivity | % Efficiency | % Predicted Hit Reliability | % Predicted No-Hit Reliability | % Overall Reliability |
| **Low Toxicity Thresholds**[a] | | | | | | | | |
| *Chironomus* survival | 51 | 36 | 8 | 52 | 84 | 53 | 93 | 79 |
| *Chironomus* biomass | 82 | 29 | 16 | 65 | 76 | 47 | 92 | 74 |
| *Hyalella* survival | 40 | 33 | 4 | 57 | 90 | 55 | 97 | 88 |
| *Hyalella* biomass | 63 | 39 | 12 | 36 | 73 | 50 | 85 | 64 |
| **High Toxicity Thresholds**[b] | | | | | | | | |
| *Chironomus* survival | 45 | 34 | 6 | 56 | 90 | 53 | 96 | 86 |
| *Chironomus* biomass | 55 | 26 | 12 | 57 | 81 | 45 | 95 | 78 |
| *Hyalella* survival[c] | 5 | 47 | 1 | 53 | 98 | 83 | 97 | 95 |
| *Hyalella* biomass | 50 | 59 | 6 | 38 | 87 | 52 | 91 | 81 |

[a] Predictions are based on a reduced chemical list that excluded petroleum compounds, unreliable models, and conventionals.

[b] Predictions are based on a reduced chemical list that excluded petroleum compounds, unreliable models, conventionals, and chemicals with high false positive error rates within each toxicity endpoint.

[c] Only five chemicals were left in this LRM set: 2,4'-DDE, sum DDEs, beta-HCH, carbazole, and total chlordane. As a result, toxicity predictions were not made for six stations because of missing values for all five of these chemicals.

HCH – hexachlorocyclohexane

LRM – logistic regression model

# 2.0 REFERENCES

Avocet. 2003. Development of freshwater sediment quality values for use in Washington State. Phase II report: Development and recommendation of SQVs for freshwater sediments in Washington State. Publication No. 03-09-088. Prepared for Washington Department of Ecology. Avocet Consulting, Kenmore, WA.

Field LJ, MacDonald DD, Norton SB, Ingersoll CG, Severn CG, Smorong D, Lindskoog R. 2002. Predicting amphipod toxicity from sediment chemistry using logistic regression models. Environ Toxicol Chem 21(9):1993–2005.

Integral, Windward, Kennedy/Jenks, Anchor. 2007. Portland Harbor RI/FS: Comprehensive round 2 site characterization summary and data gaps analysis report, plus addenda. IC07-0004. Prepared for Lower Willamette Group. Integral Consulting, Inc., Mercer Island, WA; Windward Environmental LLC, Seattle, WA; Kennedy/Jenks Consultants, Portland, OR; Anchor Environmental, LLC, Portland, OR.

Windward. 2009. Portland Harbor RI, Appendix G: Baseline ecological risk assessment. Draft. Prepared for the Lower Willamette Group. August 19, 2009. Windward Environmental LLC, Seattle, WA.

PORTLAND HARBOR RI

# BENTHIC TOXICITY REANALYSIS TECHNICAL MEMORANDUM

## APPENDIX D
## GENERIC SQG RELIABILITY ANALYSIS

**DRAFT**

**DO NOT QUOTE OR CITE**
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

November 13, 2009

**Prepared for**
The Lower Willamette Group

**Prepared by**
Windward Environmental LLC

# TABLE OF CONTENTS

# LIST OF TABLES

# 1.0 RELIABILITY ANALYSIS OF GENERIC SQGS FOR INDIVIDUAL CHEMICALS

The predictions of toxicity based on each set of high generic sediment quality guidelines (SQGs) were compared to the site-specific toxicity test results for both sets of low and high toxicity thresholds derived using the Calcasieu and the US Environmental Protection Agency (EPA) approaches. Similar to the reliability analysis for the site-specific SQGs, the acceptability criterion for the false negative and false positive rates was ≤ 20%. Tables 1-1 and 1-2 present the false negative and false positive rates for the five sets of high generic SQGs compared with the site-specific toxicity test results based on the Calcasieu approach, and Tables 1-3 and 1-4 present the false negative and false positive rates for the five sets of high generic SQGs compared with the site-specific toxicity test results based on the EPA 2009 approach. None of the five sets of high generic SQGs based on the Calcasieu or EPA 2009 approaches could reliably predict toxicity as an empirical test result exceeding either the high or low toxicity thresholds.

**Table 1-1. False Positive and False Negative Rates Associated with Each Set of High SQGs Based on a Comparison with Calcasieu High Toxicity Thresholds**

| | | ***Chironomus*** | | | | ***Hyalella*** | | | | **All Four Endpoints Combined (%)** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Mortality (%)** | | **Biomass (%)** | | **Mortality (%)** | | **Biomass (%)** | | | |
| **SQG Set** | **Toxicity Test Sampling Locations Above Generic SQG Set (%)** | **False Positive** | **False Negative** | **False Positive** | **False Negative** | **False Positive** | **False Negative** | **False Positive** | **False Negative** | **False Positive** | **False Negative** |
| PEC | 35 | 32 | 29 | 30 | 22 | 32 | 17 | 33 | 19 | 30 | 26 |
| PEL | 54 | 50 | 14 | 49 | 11 | 51 | 5.6 | 51 | 6.3 | 49 | 13 |
| RSET SL2 | 38 | 35 | 32 | 33 | 28 | 35 | 22 | 36 | 25 | 33 | 29 |
| ERM | 41 | 37 | 25 | 36 | 22 | 38 | 17 | 39 | 19 | 36 | 24 |
| CSL | 28 | 25 | 43 | 23 | 36 | 26 | 33 | 26 | 31 | 23 | 37 |

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

**Table 1-2. False Positive and Negative Rates Associated with Each Set of High SQGs Based on a Comparison with Calcasieu Low Toxicity Thresholds**

| SQG Set | Toxicity Test Sampling Locations Above Generic SQG Set (%) | *Chironomus* Mortality (%) | | *Chironomus* Biomass (%) | | *Hyalella* Mortality (%) | | *Hyalella* Biomass (%) | | All Four Endpoints Combined (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative |
| PEC | 35 | 31 | 35 | 29 | 27 | 32 | 24 | 33 | 40 | 30 | 42 |
| PEL | 54 | 49 | 16 | 48 | 16 | 51 | 9.5 | 51 | 20 | 48 | 25 |
| RSET SL2 | 38 | 34 | 38 | 33 | 32 | 35 | 29 | 36 | 43 | 33 | 44 |
| ERM | 41 | 37 | 32 | 35 | 25 | 38 | 24 | 39 | 43 | 36 | 41 |
| CSL | 28 | 25 | 46 | 22 | 39 | 26 | 38 | 26 | 53 | 23 | 51 |

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

**Table 1-3. False Positive and False Negative Rates Associated with Each Set of High SQGs Based on a Comparison with EPA 2009 High Toxicity Thresholds**

| SQG Set | Toxicity Test Sampling Locations Above Generic SQG Set (%) | *Chironomus* Mortality (%) False Positive | *Chironomus* Mortality (%) False Negative | *Chironomus* Biomass (%) False Positive | *Chironomus* Biomass (%) False Negative | *Hyalella* Mortality (%) False Positive | *Hyalella* Mortality (%) False Negative | *Hyalella* Biomass (%) False Positive | *Hyalella* Biomass (%) False Negative | All Four Endpoints Combined (%) False Positive | All Four Endpoints Combined (%) False Negative |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PEC | 35 | 31 | 31 | 29 | 26 | 32 | 21 | 33 | 51 | 30 | 46 |
| PEL | 54 | 50 | 16 | 48 | 14 | 51 | 5.3 | 51 | 31 | 49 | 29 |
| RSET SL2 | 38 | 34 | 34 | 33 | 33 | 35 | 26 | 37 | 54 | 34 | 49 |
| ERM | 41 | 37 | 28 | 35 | 26 | 38 | 21 | 40 | 54 | 37 | 46 |
| CSL | 28 | 25 | 44 | 23 | 40 | 26 | 37 | 27 | 62 | 24 | 55 |

CSL – cleanup screening level
EPA – US Environmental Protection Agency
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

**Table 1-4. False Positive and Negative Rates Associated with Each Set of High SQGs Based on a Comparison with EPA 2009 Low Toxicity Thresholds**

| SQG Set | Toxicity Test Sampling Locations Above Generic SQG Set (%) | *Chironomus* Mortality (%) False Positive | *Chironomus* Mortality (%) False Negative | *Chironomus* Biomass (%) False Positive | *Chironomus* Biomass (%) False Negative | *Hyalella* Mortality (%) False Positive | *Hyalella* Mortality (%) False Negative | *Hyalella* Biomass (%) False Positive | *Hyalella* Biomass (%) False Negative | All Four Endpoints Combined (%) False Positive | All Four Endpoints Combined (%) False Negative |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PEC | 35 | 31 | 41 | 28 | 31 | 32 | 24 | 36 | 67 | 32 | 59 |
| PEL | 54 | 49 | 23 | 47 | 15 | 51 | 9.5 | 54 | 48 | 51 | 43 |
| RSET SL2 | 38 | 34 | 39 | 32 | 33 | 35 | 29 | 39 | 64 | 35 | 57 |
| ERM | 41 | 37 | 34 | 34 | 25 | 38 | 24 | 44 | 68 | 40 | 57 |
| CSL | 28 | 24 | 50 | 21 | 38 | 26 | 38 | 30 | 76 | 26 | 68 |

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

# 2.0 RELIABILITY ANALYSIS OF MEAN QUOTIENTS

The predictive power of the mean quotients was assessed by evaluating the false positive and false negative rates; specifically, empirical toxicity test results based on the Calcasieu and EPA 2009 approaches were compared to the mean quotient predictions of toxicity using the thresholds of 1.0 and 0.7 as stated in EPA's problem formulation as presented in the draft baseline ecological risk assessment (BERA) (Windward 2009). Tables 2-1 and 2-2 present the false negative and false positive rates for the mean quotients based on the five sets of high generic SQGs for the Calcasieu high and low thresholds, respectively; Tables 2-3 and 2-4 present the false negative and false positive rates for the mean quotients based on the five sets of high generic SQGs for the EPA 2009 high and low thresholds, respectively. The percentage of sampling locations predicted to have adverse effects on benthic invertebrates based on the mean quotients ranged from 6 to 15%. Compared with the toxicity test results based on the high toxicity thresholds for the Calcasieu and EPA 2009 approaches, all mean quotients for the separate endpoints had low false positive rates (3.1 to 12%) and high false negative rates (28 to 82%) (Tables 2-1 and 2-3). Combining the toxicity test results based on all four endpoints (adverse effects in any endpoint was counted as adverse effect for that sample) gave similar low false positive rates (3.1 to 9.0%) and high false negative rates (47 to 83%). The false positive and false negative rates were similar for the low toxicity threshold comparison (false positive rates of 2.9 to 13% and false negative rates of 33 to 91%) for all mean quotients of the separate endpoints based on the Calcasieu and EPA 2009 approaches (Tables 2-2 and 2-4). Combining the toxicity test results for all four endpoints gave similar low false positive rates (3.4 to 9.2%) and high false negative rates (63 to 89%).

**Table 2-1. False Positive and Negative Rates Associated with Each High SQG Quotient Based on a Comparison with Calcasieu High Toxicity Thresholds**

| SQGs Used to Calculate Mean Quotients | Toxicity Test Sampling Locations Above Threshold (%) | *Chironomus* Mortality (%) | | *Chironomus* Biomass (%) | | *Hyalella* Mortality (%) | | *Hyalella* Biomass (%) | | All Four Endpoints Combined (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative |
| PEC | 10 | 6.4 | 54 | 5.1 | 53 | 6.2 | 28 | 6.9 | 31 | 5.1 | 55 |
| PEL | 15 | 11 | 46 | 8.9 | 44 | 11 | 33 | 12 | 31 | 9.0 | 47 |
| RSET SL2 | 9 | 6.0 | 64 | 4.7 | 61 | 6.2 | 50 | 6.1 | 44 | 4.7 | 63 |
| ERM | 11 | 7.5 | 54 | 7.4 | 61 | 8.4 | 44 | 9.0 | 50 | 7.5 | 63 |
| CSL | 6 | 3.8 | 68 | 3.1 | 69 | 4.4 | 61 | 4.3 | 56 | 3.1 | 71 |

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

**Table 2-2. False Positive and Negative Rates Associated with Each High SQG Quotient Based on a Comparison with Calcasieu Low Toxicity Thresholds**

| SQGs Used to Calculate Mean Quotients | Toxicity Test Sampling Locations Above Threshold (%) | *Chironomus* Mortality (%) | | *Chironomus* Biomass (%) | | *Hyalella* Mortality (%) | | *Hyalella* Biomass (%) | | All Four Endpoints Combined (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative |
| PEC | 10 | 5.5 | 57 | 4.4 | 57 | 5.9 | 33 | 6.1 | 53 | 4.7 | 68 |
| PEL | 15 | 9.8 | 51 | 8.4 | 50 | 11 | 38 | 11 | 57 | 9.0 | 63 |
| RSET SL2 | 9 | 5.1 | 65 | 4.0 | 64 | 5.9 | 52 | 6.1 | 67 | 4.3 | 73 |
| ERM | 11 | 7.8 | 65 | 7.2 | 66 | 8.5 | 52 | 8.7 | 67 | 7.7 | 75 |
| CSL | 6 | 3.1 | 70 | 3.2 | 75 | 4.4 | 67 | 4.6 | 77 | 3.4 | 81 |

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

**Table 2-3. False Positive and Negative Rates Associated with Each High SQG Quotient Based on a Comparison with EPA 2009 High Toxicity Thresholds**

| SQGs Used to Calculate Mean Quotients | Toxicity Test Sampling Locations Above Threshold (%) | *Chironomus* Mortality (%) | | *Chironomus* Biomass (%) | | *Hyalella* Mortality (%) | | *Hyalella* Biomass (%) | | All Four Endpoints Combined (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative |
| PEC | 10 | 5.7 | 53 | 4.8 | 57 | 6.2 | 32 | 6.3 | 64 | 4.8 | 71 |
| PEL | 15 | 10 | 47 | 8.8 | 50 | 11 | 37 | 11 | 64 | 8.8 | 65 |
| RSET SL2 | 9 | 5.4 | 63 | 4.0 | 62 | 6.2 | 53 | 6.3 | 74 | 4.4 | 75 |
| ERM | 11 | 7.7 | 59 | 7.6 | 67 | 8.4 | 47 | 9.1 | 74 | 7.9 | 77 |
| CSL | 6 | 3.1 | 66 | 3.2 | 74 | 4.4 | 63 | 4.7 | 82 | 3.5 | 83 |

CSL – cleanup screening level
EPA – US Environmental Protection Agency
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

**Table 2-4. False Positive and Negative Rates Associated with Each High SQG Quotient Based on a Comparison with EPA 2009 Low Toxicity Thresholds**

| SQGs Used to Calculate Mean Quotients | Toxicity Test Sampling Locations Above Threshold (%) | *Chironomus* Mortality (%) | | *Chironomus* Biomass (%) | | *Hyalella* Mortality (%) | | *Hyalella* Biomass (%) | | All Four Endpoints Combined (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative | False Positive | False Negative |
| PEC | 10 | 5.6 | 64 | 4.1 | 62 | 5.9 | 33 | 7.3 | 81 | 5.4 | 81 |
| PEL | 15 | 9.6 | 57 | 7.5 | 52 | 11 | 38 | 13 | 81 | 9.2 | 76 |
| RSET SL2 | 9 | 5.2 | 70 | 3.7 | 67 | 5.9 | 52 | 7.3 | 87 | 4.9 | 84 |
| ERM | 11 | 8.0 | 70 | 7.1 | 69 | 8.5 | 52 | 10 | 85 | 8.6 | 84 |
| CSL | 6 | 3.2 | 75 | 2.9 | 77 | 4.4 | 67 | 5.5 | 91 | 3.8 | 89 |

CSL – cleanup screening level
EPA – US Environmental Protection Agency
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

Because none of the mean quotients that used the thresholds of 1.0 or 0.7 met the acceptability criterion of ≤ 20% false negative rates, the use of lower mean quotient thresholds was evaluated. Similar to the approach used in other studies (Long et al. 2006), the mean quotient thresholds were adjusted in an attempt to achieve acceptable error rates in the comparison with toxicity test results based on the high and low toxicity thresholds for the Calcasieu and EPA 2009 approaches. Because the acceptability criterion for both false negative and false positive rates for both approaches could not be met, even with adjustments to the quotient threshold, the high SQG mean quotients were not used in the benthic risk characterization.

# 3.0 RELIABILITY ANALYSIS OF ESBS

Equilibrium partitioning sediment benchmarks (ESBs) were calculated for polycyclic aromatic hydrocarbon (PAH) mixtures, non-ionic organic compounds, and select pesticides. The predictive power of the ESB for PAH mixtures was assessed by evaluating the false positive and false negative rates based on a comparison of the empirical toxicity test results for the Calcasieu and EPA 2009 approaches and the ESB predictions of toxicity. Table 3-1 presents the false negative and false positive rates for the ESB for PAH mixtures, gamma-hexachlorocyclohexane (HCH), and the sum of PAHs and narcotic non-ionic organics for the Calcasieu high and low thresholds; Table 3-2 presents the false negative and false positive rates for the ESB for PAH mixtures, gamma-HCH, and the sum of PAHs and narcotic non-ionic organics for the EPA 2009 high and low thresholds. The ESB for PAH mixtures predicted that 18% of the sampling locations had adverse effects on benthic invertebrates. All endpoints had low false positive rates (11 to 18%) and high false negative rates (41 to 83%). Combining the toxicity test results across all four endpoints (adverse effects in any endpoint was counted as adverse effect for that sample) had only slight effects on the false positive and negative rates. Similar to the mean quotient results, the thresholds were adjusted in an attempt to achieve acceptable error rates in the comparison with toxicity test results based on the high and low toxicity thresholds for the Calcasieu and EPA 2009 approaches. In both comparisons, the acceptability criterion of ≤ 20% false positive and false negative rates could not be met; hence, the ESB for PAH mixtures was not carried forward in the evaluation of risks to the benthic community.

**Table 3-1. False Negative and False Positive Rates Associated with ESBs Based on Calcasieu High and Low Toxicity Thresholds**

| | | *Chironomus* | | *Hyalella* | | Four Endpoints Combined |
|---|---|---|---|---|---|---|
| **ESB Group** | **Parameter** | **Mortality** | **Biomass** | **Mortality** | **Biomass** | |
| PAH mixtures | **High toxicity thresholds** | | | | | |
| | % False negative | 43 | 44 | 44 | 44 | 45 |
| | % False positive | 13 | 12 | 15 | 15 | 12 |
| | **Low toxicity thresholds** | | | | | |
| | % False negative | 46 | 52 | 43 | 63 | 63 |
| | % False positive | 12 | 12 | 15 | 16 | 13 |
| gamma-HCH | **High toxicity thresholds** | | | | | |
| | % False negative | 80 | 86 | 67 | 50 | 86 |
| | % False positive | 23 | 24 | 22 | 21 | 24 |
| | **Low toxicity thresholds** | | | | | |
| | % False negative | 83 | 89 | 67 | 80 | 91 |
| | % False positive | 23 | 24 | 22 | 23 | 26 |

**Table 3-1. False Negative and False Positive Rates Associated with ESBs Based on Calcasieu High and Low Toxicity Thresholds**

| ESB Group | Parameter | *Chironomus* Mortality | *Chironomus* Biomass | *Hyalella* Mortality | *Hyalella* Biomass | Four Endpoints Combined |
|---|---|---|---|---|---|---|
| Sum of PAHs and narcotic nonionic organics | **High toxicity thresholds** | | | | | |
| | % False negative | 43 | 44 | 44 | 44 | 45 |
| | % False positive | 13 | 12 | 15 | 15 | 12 |
| | **Low toxicity thresholds** | | | | | |
| | % False negative | 46 | 52 | 43 | 63 | 63 |
| | % False positive | 12 | 12 | 15 | 16 | 13 |

ESB – equilibrium partitioning sediment benchmark
HCH – hexachlorocyclohexane
PAH – polycyclic aromatic hydrocarbon

**Table 3-2. False Negative and False Positive Rates Associated with ESB Based on EPA 2009 High and Low Toxicity Thresholds**

| ESB Group | Parameter | *Chironomus* Mortality | *Chironomus* Biomass | *Hyalella* Mortality | *Hyalella* Biomass | Four Endpoints Combined |
|---|---|---|---|---|---|---|
| PAH mixtures | **High toxicity thresholds** | | | | | |
| | % False negative | 41 | 50 | 42 | 69 | 65 |
| | % False positive | 13 | 12 | 15 | 16 | 13 |
| | **Low toxicity thresholds** | | | | | |
| | % False negative | 52 | 54 | 43 | 83 | 77 |
| | % False positive | 12 | 11 | 15 | 18 | 14 |
| gamma-HCH | **High toxicity thresholds** | | | | | |
| | % False negative | 80 | 88 | 67 | 83 | 92 |
| | % False positive | 23 | 24 | 22 | 23 | 26 |
| | **Low toxicity thresholds** | | | | | |
| | % False negative | 88 | 90 | 67 | 92 | 95 |
| | % False positive | 24 | 25 | 22 | 26 | 31 |
| Sum of PAHs and narcotic nonionic organics | **High toxicity thresholds** | | | | | |
| | % False negative | 41 | 50 | 42 | 69 | 65 |
| | % False positive | 13 | 12 | 15 | 16 | 13 |
| | **Low toxicity thresholds** | | | | | |
| | % False negative | 52 | 54 | 43 | 83 | 77 |
| | % False positive | 12 | 11 | 15 | 18 | 14 |

EPA – US Environmental Protection Agency
ESB – equilibrium partitioning sediment benchmark
HCH – hexachlorocyclohexane
PAH – polycyclic aromatic hydrocarbon

For the non-ionic organic chemicals, a reliability analysis was only performed for gamma-HCH and the sum of PAHs and narcotic non-ionic organics because all the other non-ionic organic chemicals had low exceedance frequencies (< 5%), were measured only in a limited number of sediment samples, or both. The ESB for gamma-HCH predicted that 22% of the sampling locations had adverse effects on benthic invertebrates. All endpoints had relatively low false positive rates (21 to 26%) and high false negative rates (50 to 92%). Pooling the toxicity test results across all four endpoints increased the rates slightly. Similar to the reliability analysis of the ESB for PAH mixtures, the thresholds were adjusted in an attempt to achieve acceptable error rates in the comparison with toxicity test results based on the high and low toxicity thresholds for the Calcasieu and EPA 2009 approaches. In all comparisons, the acceptability criterion of ≤ 20% false positive and false negative rates could not be met.

The ESB for the sum of PAHs and narcotic non-ionic organics predicted that 18% of the sampling locations had adverse effects on benthic invertebrates. The false positive and false negative rates were the same as those for the ESB PAH mixture. Similar to the reliability analysis of the ESB for PAH mixtures, the thresholds were adjusted in an attempt to achieve acceptable error rates in the comparison with toxicity test results based on the high and low toxicity thresholds for the Calcasieu and EPA 2009 approaches. In all comparisons, the acceptability criterion of ≤ 20% false positive and false negative rates could not be met. Because the ESB for gamma-HCH and the sum of PAHs and narcotic non-ionic organics did not meet the acceptability criterion for the false positive and false negative rates, these SQGs were not carried forward in the evaluation of risks to the benthic community.

# 4.0 RELIABILITY ANALYSIS OF LOW SQGS

The ability of the sets of low SQGs to predict the absence of toxicity was evaluated based on a comparison of the SQG exceedances to the low toxicity thresholds derived in the Calcasieu and EPA 2009 approaches. Low SQGs included threshold effects concentration (TEC), threshold effects level (TEL), sediment quality standards, lowest apparent effects threshold, effects range – low, and screening level 1 thresholds. Because the narrative intent of the low SQGs is to identify the concentrations of chemicals of potential concern (COPCs) or COPC mixtures below which adverse effects on benthic invertebrates would be infrequently observed, the analysis of the low SQGs focused on false negative rates for each set of low SQGs. The false negative rate is the number of toxicity test samples that were predicted to be non-toxic by the SQGs but were toxic based on the toxicity test results divided by the sum of correctly predicted toxic locations and the incorrectly predicted non-toxic locations (i.e., all those samples that were actually toxic). Hence, this error rate identifies the proportion of samples that is erroneously deemed as being clean. Similar to the reliability analysis of the high SQGs, an acceptability criterion of ≤ 20% was selected for false negative rates.

Table 4-1 presents the false negative rates associated with each set of low SQGs based on the Calcasieu and EPA 2009 approaches. Three sets of low SQGs (TECs, TELs, and ERLs) met the acceptability criterion of a ≤ 20% false negative rate.

**Table 4-1. Number of Stations Predicted to be Non-Toxic Based on the Sets of Low SQGs and False Negative Rates for the Calcasieu and EPA 2009 Approaches**

| | | Calcasieu Approach | | | | EPA 2009 Approach | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Percentage of Toxicity Test Stations Below Threshold | *Chironomus* | | *Hyalella* | | *Chironomus* | | *Hyalella* | |
| **SQG Set** | | **Survival (%)** | **Biomass (%)** | **Survival (%)** | **Biomass (%)** | **Survival (%)** | **Biomass (%)** | **Survival (%)** | **Biomass (%)** |
| TEC | 11 | 2.7 | 0 | 4.8 | 0 | 4.5 | 0 | 4.8 | 4.0 |
| TEL | 3 | 0 | 0 | 4.8 | 0 | 2.3 | 0 | 4.8 | 0 |
| RSET SL1 | 42 | 22 | 14 | 24 | 27 | 25 | 15 | 24 | 48 |
| ERL | 6 | 2.7 | 0 | 4.8 | 0 | 4.5 | 0 | 4.8 | 0 |
| SQS | 56 | 30 | 25 | 29 | 43 | 32 | 23 | 29 | 68 |
| LAET | 55 | 32 | 23 | 24 | 40 | 34 | 23 | 24 | 63 |

EPA – US Environmental Protection Agency
ERL – effects range – low
LAET – lowest apparent effects threshold
RSET – Regional Sediment Evaluation Team
SL1 – screening level 1
SQG – sediment quality guideline
SQS – sediment quality standards
TEC – threshold effects concentration
TEL – threshold effects level

# 5.0 REFERENCES

Ecology. 1995. Sediment management standards. WAC 173-204. Washington State Department of Ecology, Toxics Cleanup Program, Sediment Management Unit, Olympia, WA.

Gries TH, Waldow KH. 1996. Progress re-evaluating Puget Sound apparent effects thresholds (AETs). Volume I: 1994 amphipod and echinoderm larval AETs. Draft report. Prepared for Puget Sound Dredged Disposal Analysis (PSDDA) agencies. Washington Department of Ecology, Olympia, WA.

Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. Environ Manage 19:81-97.

MacDonald DD, Ingersoll CG, Berger TA. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39(5):20-31.

RSET. 2006. Regional sediment evaluation framework for the Pacific Northwest. Interim final. September 2006. Regional Sediment Evaluation Team (US Army Corps of Engineers - Seattle District, Portland District, Walla Walla District, and Northwestern Division; US Environmental Protection Agency Region 10; Washington Department of Ecology; Washington Department of Natural Resources; Oregon Department of Environmental Quality; Idaho Department of Environmental Quality; National Marine Fisheries Service; US Fish and Wildlife Service).

Smith SL, MacDonald DD, Keenleyside KA, Ingersoll CG, Field LJ. 1996. A preliminary evaluation of sediment quality assessment values for freshwater ecosystems. J Great Lakes Res 22:624-638.

PORTLAND HARBOR RI/FS

# BENTHIC TOXICITY REANALYSIS TECHNICAL MEMORANDUM

## MAP FOLIO

**DRAFT**

**DO NOT QUOTE OR CITE**
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

November 13, 2009

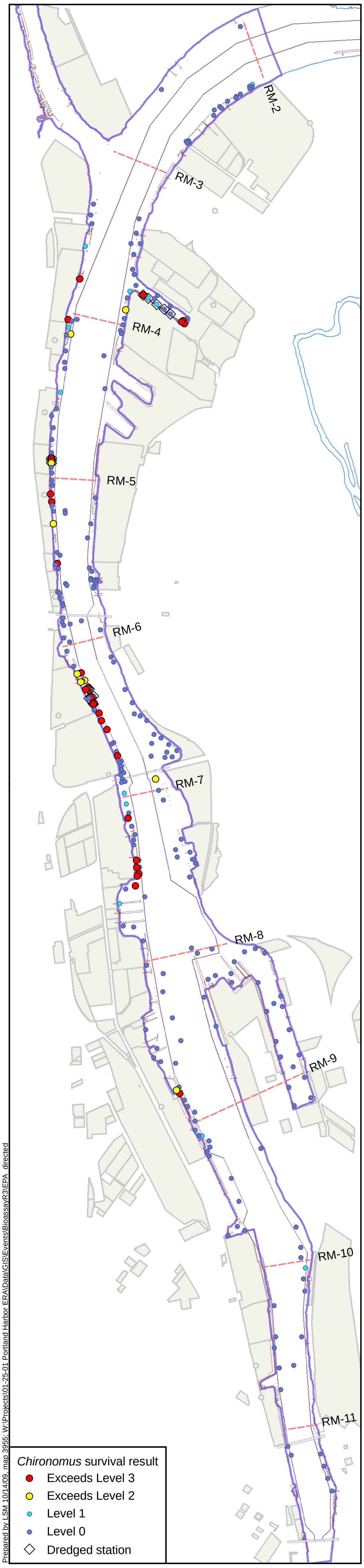
**Prepared for**
The Lower Willamette Group

**Prepared by**
Windward Environmental LLC

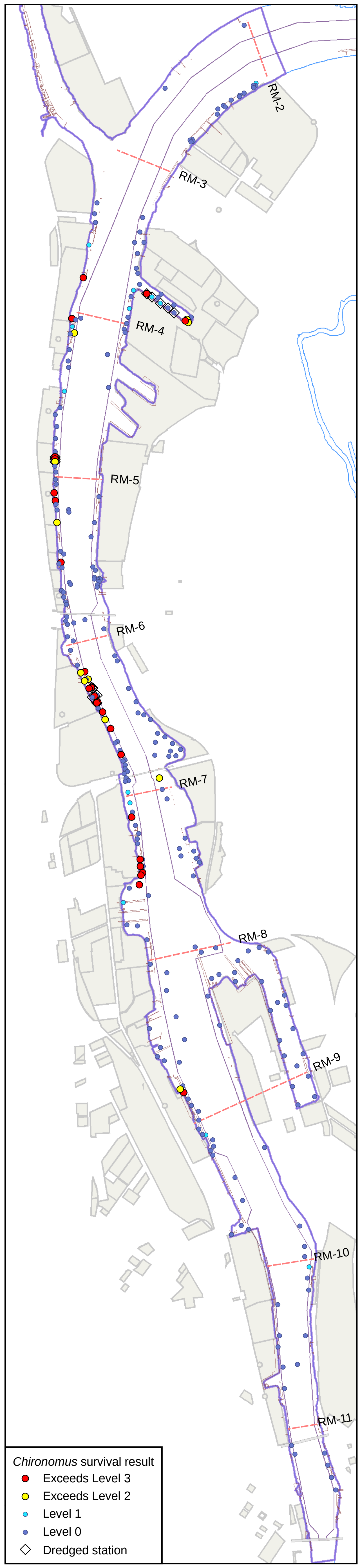
WE-09-0004

RECOMMENDED FOR INCLUSION IN ADMINISTRATIVE RECORD

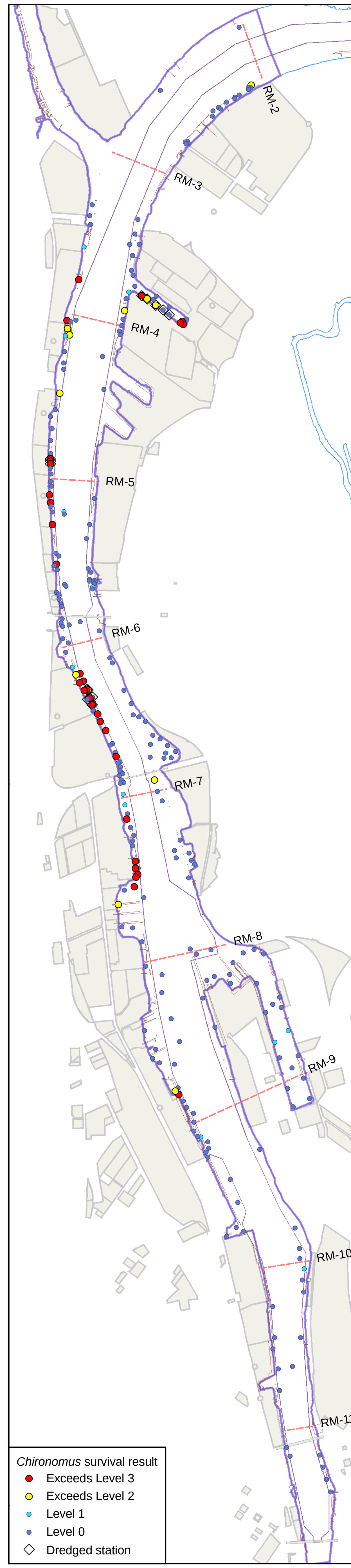
# LIST OF MAPS




**Map 3-1. *Chironomus dilutus* Survival Results Based on the Calcasieu Approach**




**Map 3-2. *Chironomus dilutus* Survival Results Based on the Draft BERA Approach**




**Map 3-3. *Chironomus dilutus* Survival Results Based on the EPA 2009 Approach**

Windward environmental LLC

LWG Lower Willamette Group

N

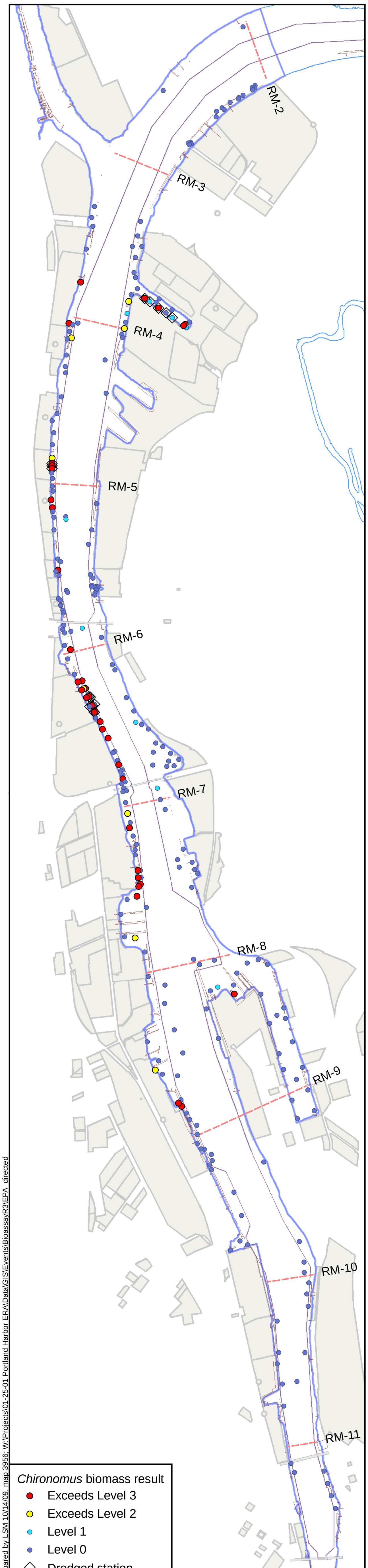
0 1,000 2,000 3,000 4,000 Feet

FEATURE SOURCES:
Transportation, Property, or Boundaries: Metro RLIS.
Channel & River miles: US Army Corps of Engineers.

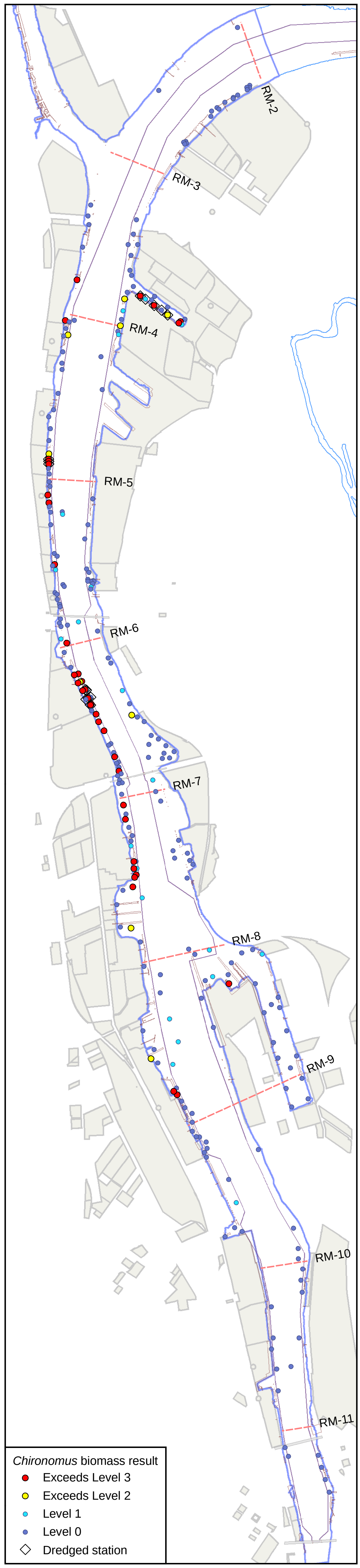
Bridge
Upland Site (2008)
Dock or Structure
River Mile
Navigation Channel
Study Area (13-foot NAVD88)
River or slough

**DRAFT**
DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

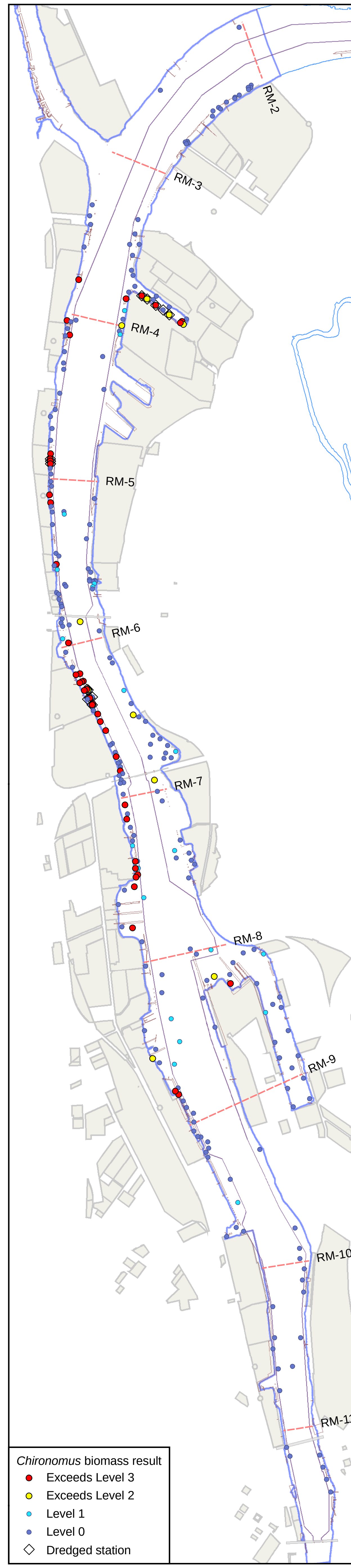
Prepared by LSM 10/14/09, map 3956; W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Events\Bioassay\R3\EPA_directed




Map 3-4. ***Chironomus dilutus*** **Biomass Results Based on the Calcasieu Approach**




Map 3-5. ***Chironomus dilutus*** **Biomass Results Based on the Draft BERA Approach**




Map 3-6. ***Chironomus dilutus*** **Biomass Results Based on the EPA 2009 Approach**

Wind Ward LLC environmental

LWG Lower Willamette Group

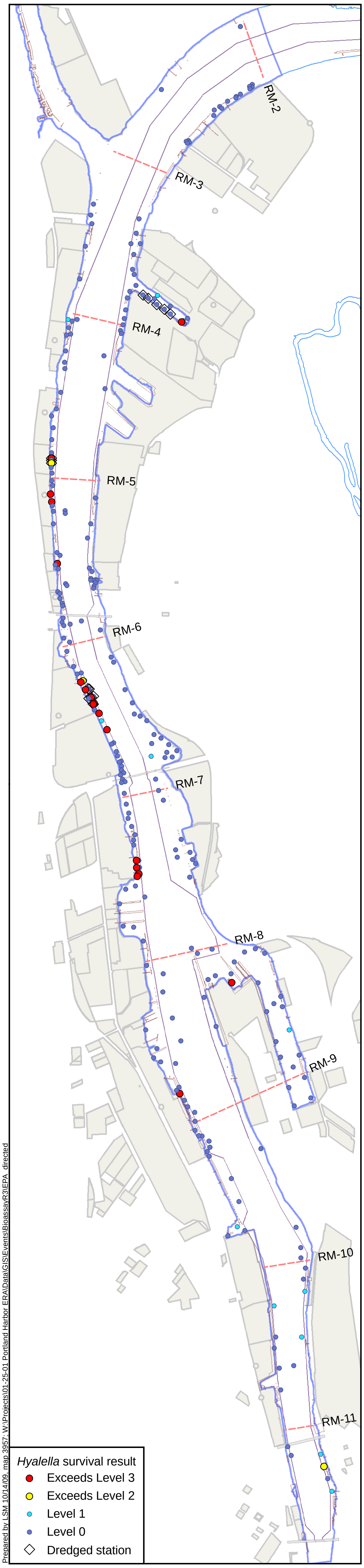
N

0 1,000 2,000 3,000 4,000 Feet

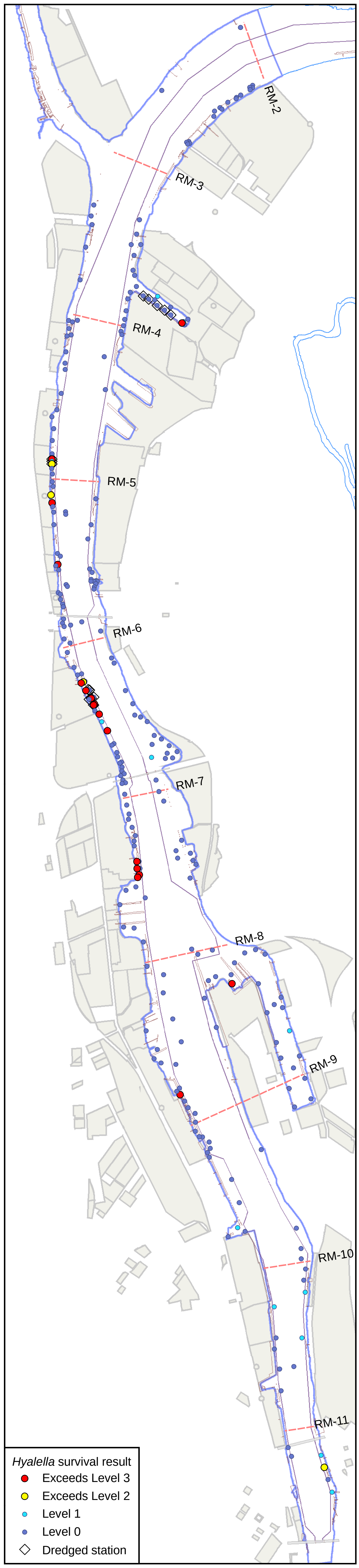
FEATURE SOURCES:
Transportation, Property, or Boundaries: Metro RLIS.
Channel & River miles: US Army Corps of Engineers.

- Bridge
- Upland Site (2008)
- Dock or Structure
- River Mile
- Navigation Channel
- Study Area (13-foot NAVD88)
- River or slough

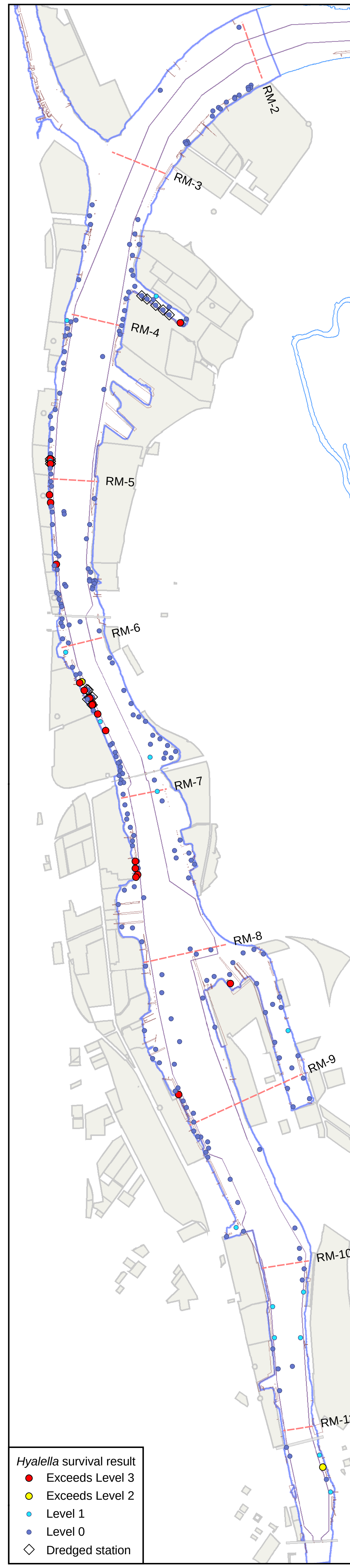
**DRAFT**
DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.




**Map 3-7. *Hyalella azteca* Survival Results Based on the Calcasieu Approach**




**Map 3-8. *Hyalella azteca* Survival Results Based on the Draft BERA Approach**




**Map 3-9. *Hyalella azteca* Survival Results Based on the EPA 2009 Approach**

Prepared by LSM 10/14/09; map 3957; W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Events\Bioassay\R3\EPA_directed

Windward environmental LLC

LWG Lower Willamette Group

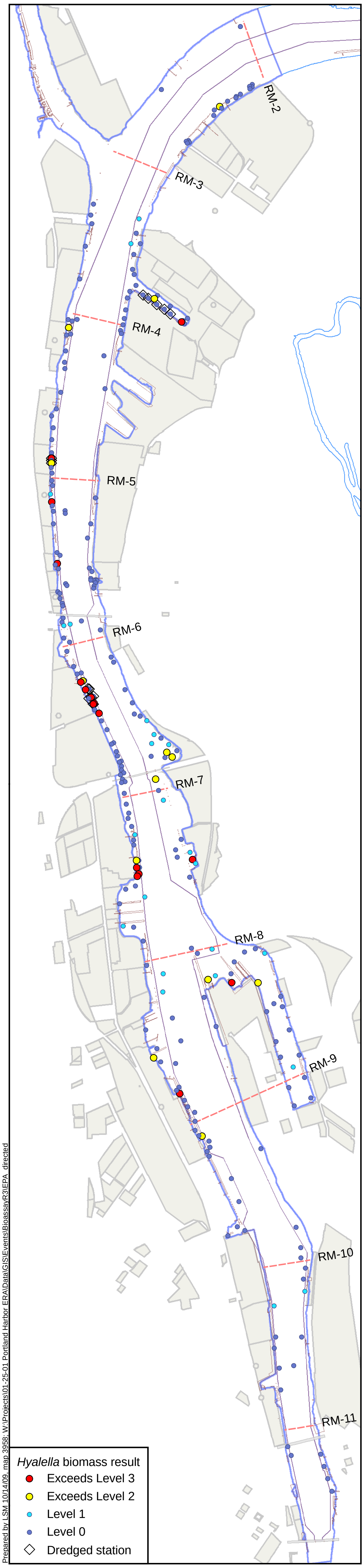
N

0 1,000 2,000 3,000 4,000 Feet

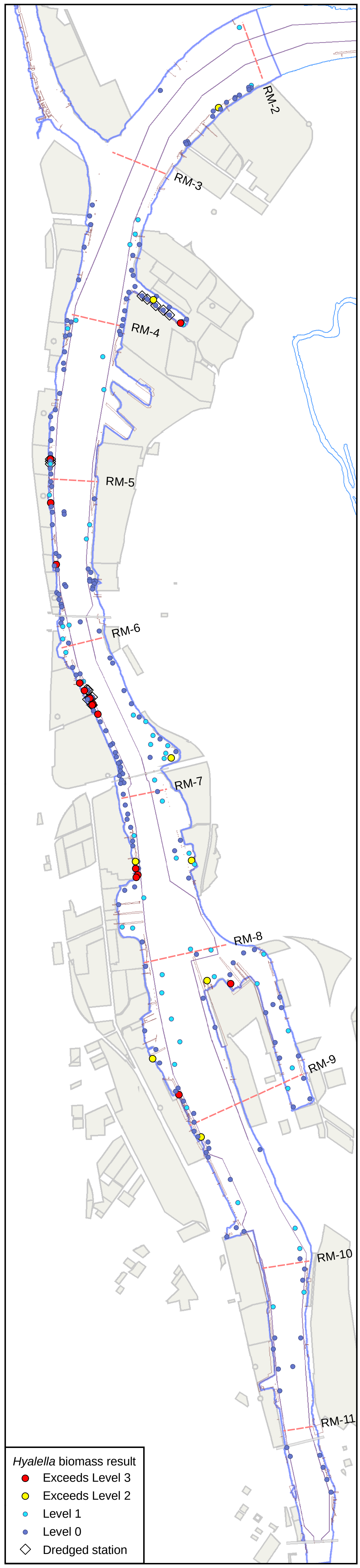
FEATURE SOURCES:
Transportation, Property, or Boundaries: Metro RLIS.
Channel & River miles: US Army Corps of Engineers.

Bridge
Upland Site (2008)
Dock or Structure
River Mile
Navigation Channel
3 Study Area (13-foot NAVD88)
River or slough

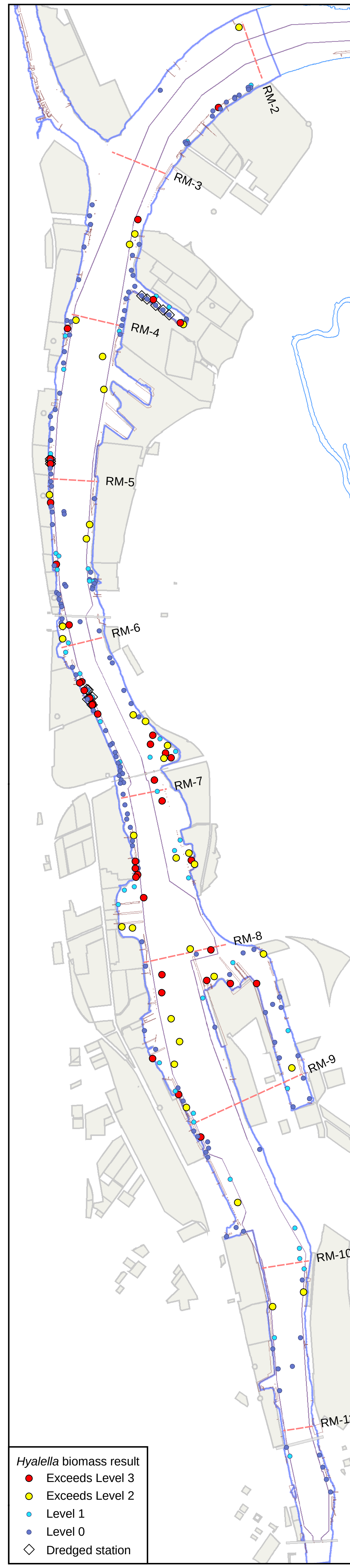
**DRAFT**
DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.




**Map 3-10. *Hyalella azteca* Bioamass Results Based on the Calcasieu Approach**




**Map 3-11. *Hyalella azteca* Biomass Results Based on the Draft BERA Approach**




**Map 3-12. *Hyalella azteca* Biomass Results Based on the EPA 2009 Approach**

Prepared by LSM 10/14/09, map 3958; W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Events\BioassayR3\EPA_directed

Windward environmental LLC

LWG Lower Willamette Group

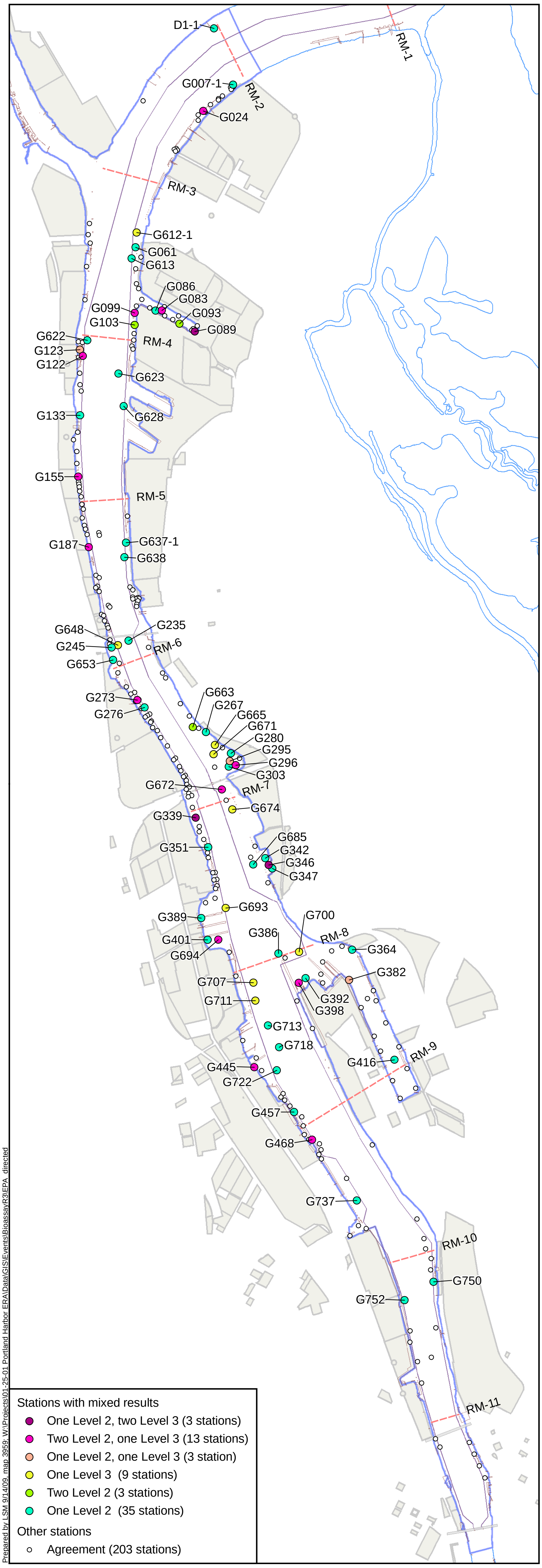
N

0 1,000 2,000 3,000 4,000 Feet

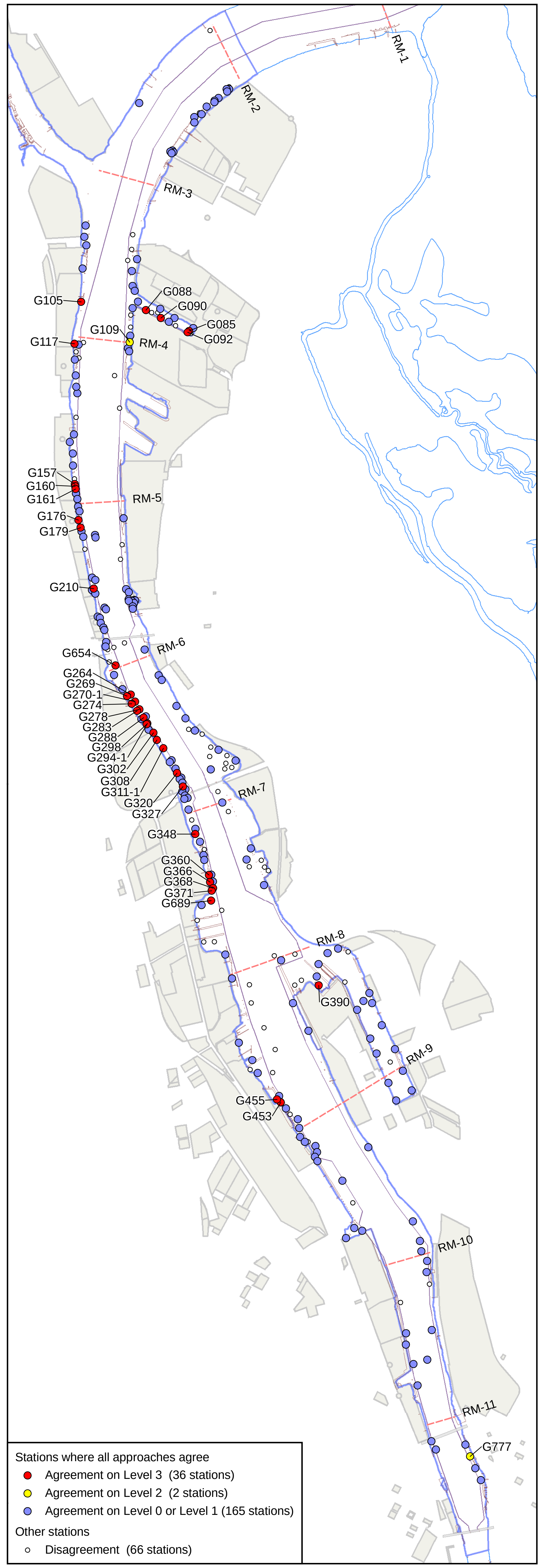
FEATURE SOURCES:
Transportation, Property, or Boundaries: Metro RLIS.
Channel & River miles: US Army Corps of Engineers.

Bridge
Upland Site (2008)
Dock or Structure
River Mile
Navigation Channel
Study Area (13-foot NAVD88)
River or slough

**DRAFT**
DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.




Map 3-13. Stations Where the Calcasieu, Draft BERA, and EPA 2009 Approaches Disagree




Map 3-14. Stations Where the Calcasieu, Draft BERA, and EPA 2009 Approaches Agree on Pooled Endpoint Results

Prepared by LSM 9/14/09, map 3959 W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Events\Bioassay\R3\EPA_directed

Windward environmental LLC

LWG Lower Willamette Group

N

0 1,000 2,000 3,000 4,000 Feet

Bridge
Upland Site (2008)
Dock or Structure
River Mile
Navigation Channel
Study Area (13-foot NAVD88)
River or slough

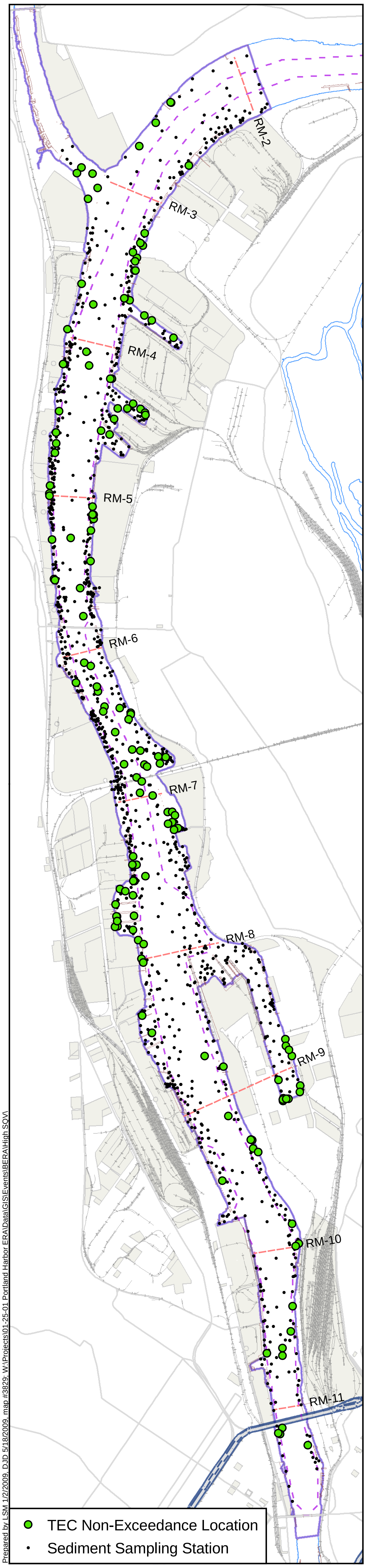
FEATURE SOURCES:
Transportation, Property, or Boundaries: Metro RLIS.
Channel & River miles: US Army Corps of Engineers.

**DRAFT**

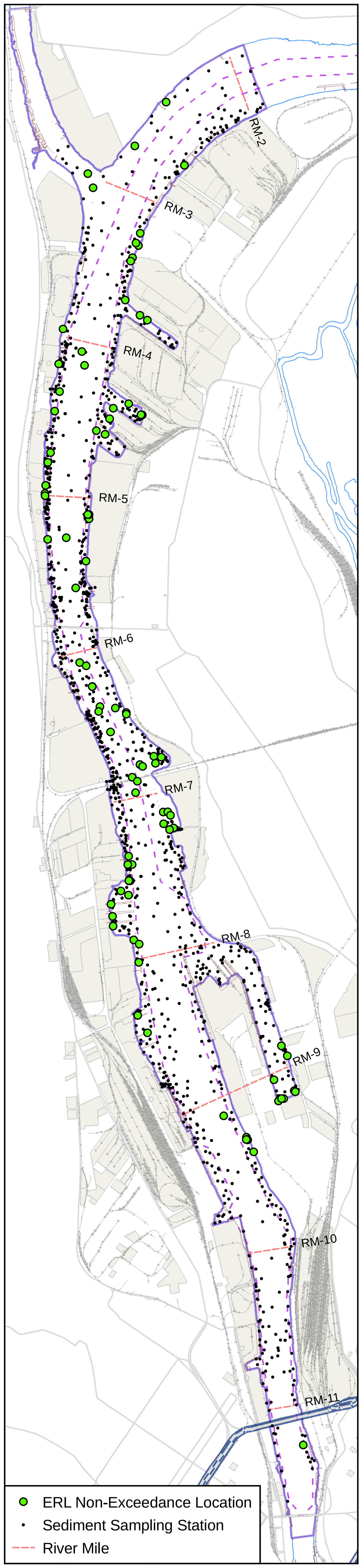
DO NOT QUOTE OR CITE

This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

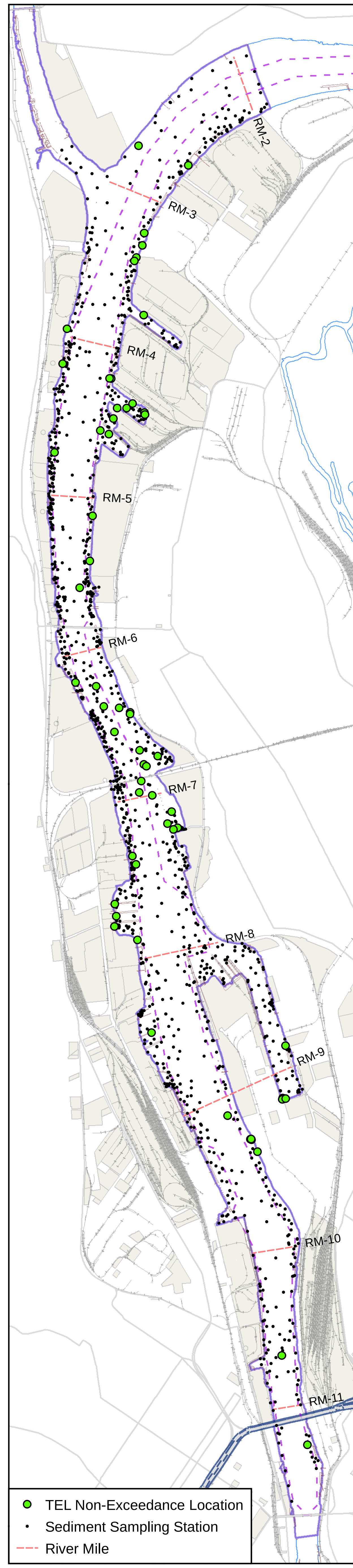
Prepared by LSM 1/2/2009, DJD 5/18/2009; map #3829; W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Events\BERA\High SQV\




**Map 5-1. Locations with No Exceedances of TEC SQGs**




**Map 5-2. Locations with No Exceedances of ERL SQGs**




**Map 5-3. Locations with No Exceedances of TEL SQGs**

Windward environmental LLC

LWG Lower Willamette Group

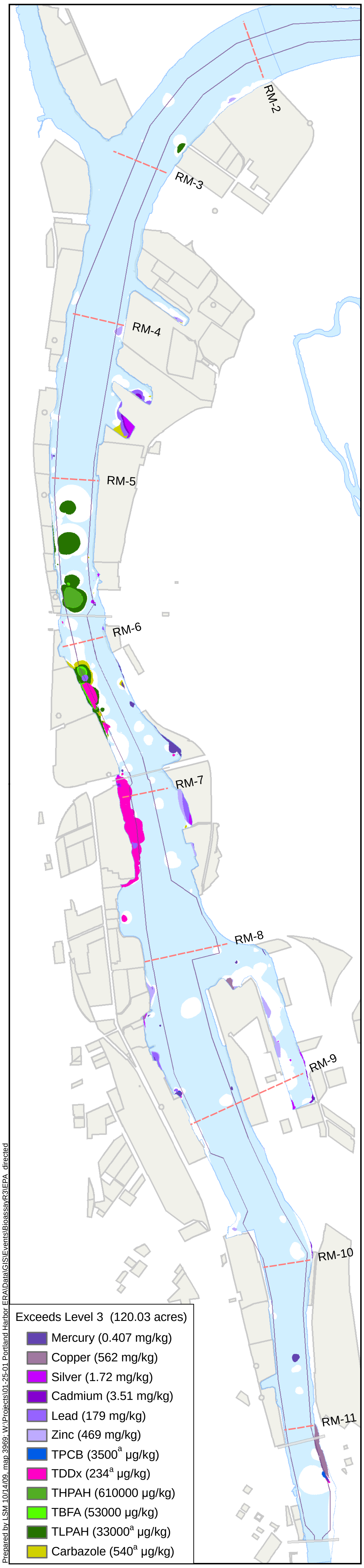
N

0 1,000 2,000 3,000 4,000 Feet

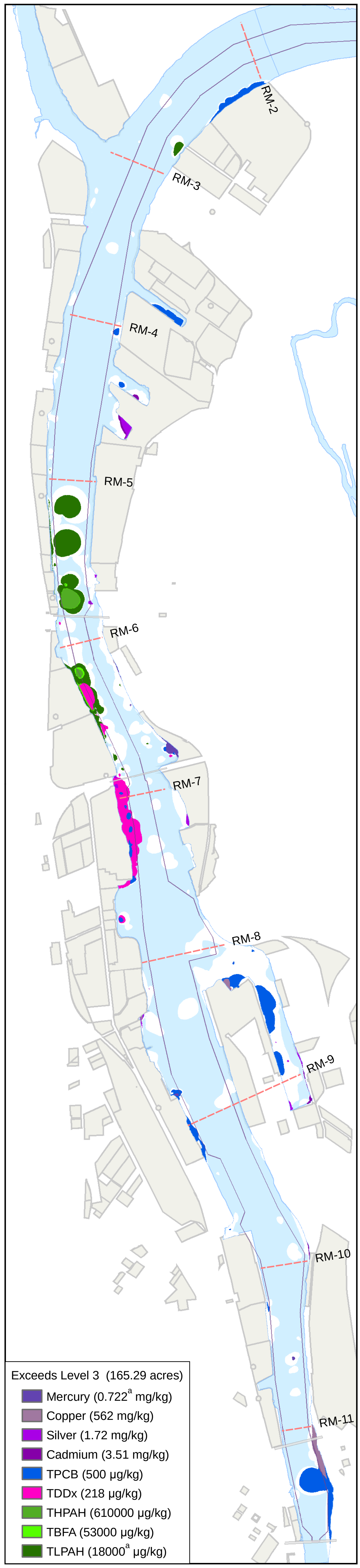
FEATURE SOURCES:
Transportation, Property, or Boundaries: Metro RLIS.
Channel & River miles: US Army Corps of Engineers.

Freeway
Arterial
Railroad
Bridge
Dock or Structure
Upland ECSI Site (2008)
Study Area
River Mile
Navigation Channel
River or Slough

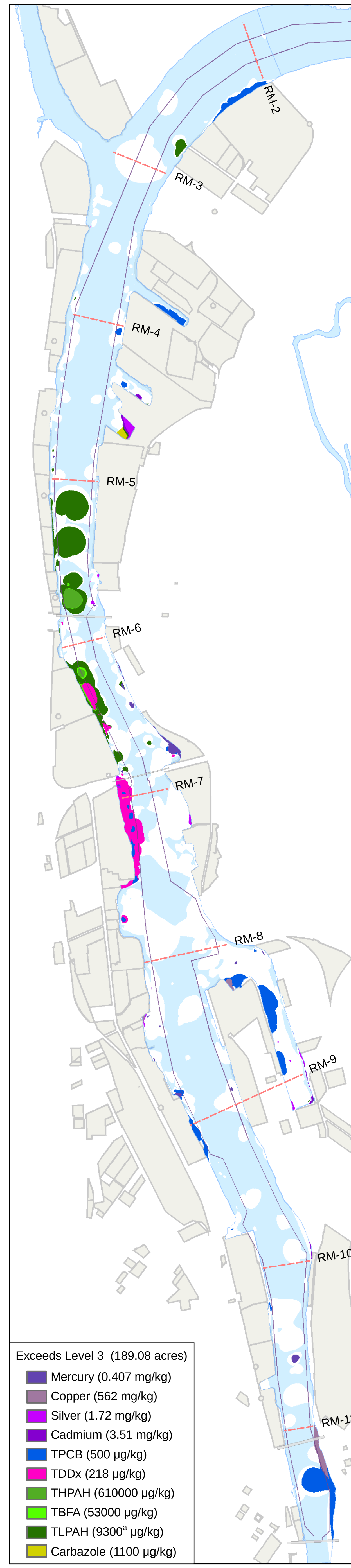
DRAFT
DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.




**Map 6-1. SQG Exceedance Areas Predicted Using the Calcasieu Approach**




**Map 6-2. SQG Exceedance Areas Predicted Using the Draft BERA Approach**




**Map 6-3. SQG Exceedance Areas Predicted Using the EPA 2009 Approach**

Prepared by LSM 10/14/09; map 3969; W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Events\Bioassay\R3\EPA_directed

Windward environmental LLC

LWG Lower Willamette Group

N

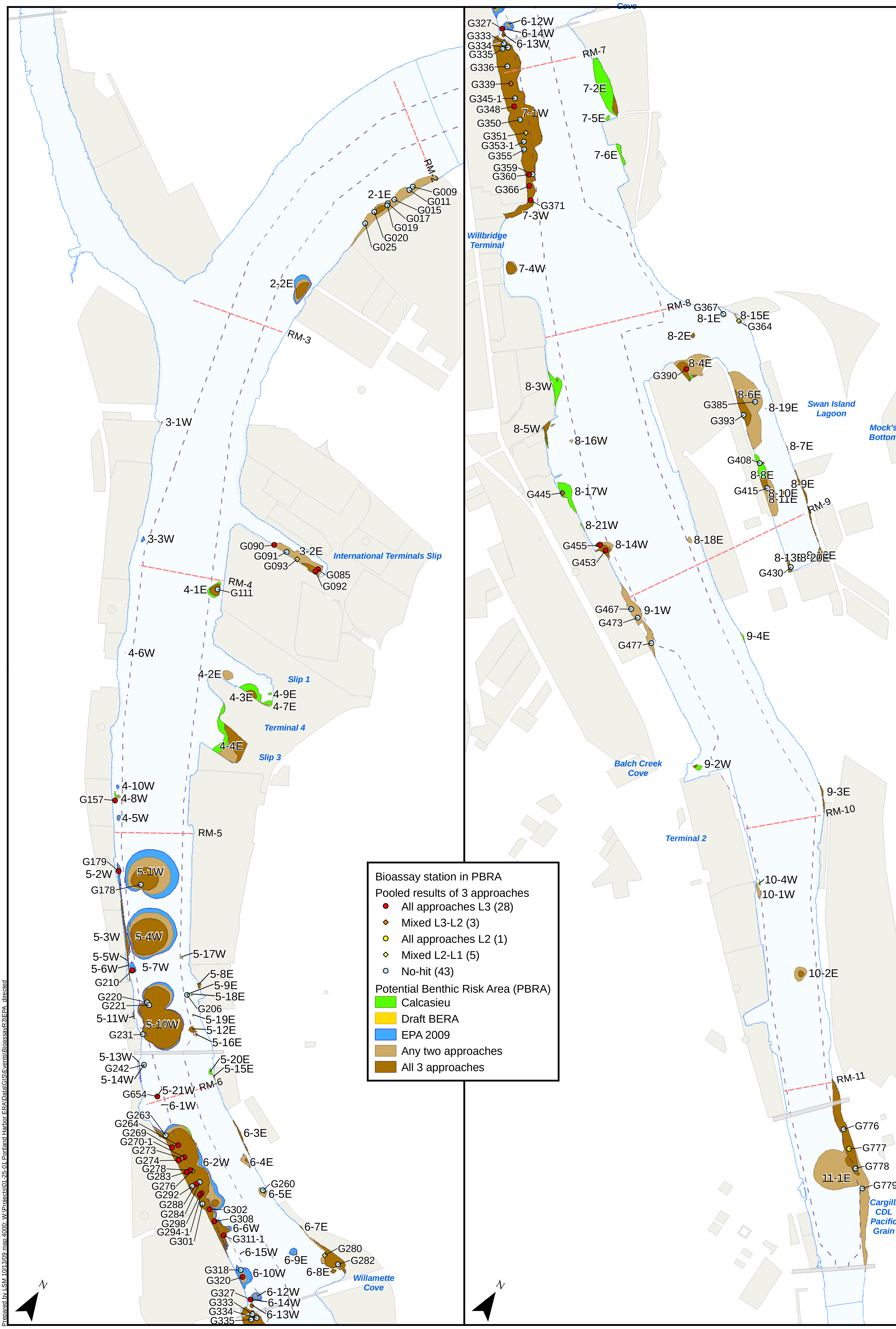
0 1,000 2,000 3,000 4,000 Feet

FEATURE SOURCES:
Transportation, Property, or Boundaries: Metro RLIS.
Channel & River miles: US Army Corps of Engineers.

Areas of Uncertainty
Clean (L0 or L1) Areas
Bridge
Upland Site (2008)
Navigation Channel
River Mile

[a]Value differs from other approaches.
Analytes with too few detected locations for reasonable interpolation:
Benzyl alcohol, 180 of 1256; dHCH, 150 of 1106; endrin, 75 of 862; endrin ketone, 187 of 1109, and dieldrin, 249 of 1104. In the draft BERA, carbazole was not interpolated.

**DRAFT**
DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.




## Map 6-4. Bioassay Stations in Potential Benthic Risk Areas from Calcasieu, Draft BERA, and EPA 2009 Approaches

Windward environmental LLC

LWG Lower Willamette Group

0 1,000 2,000 3,000 Feet

Bridge
Upland Site (2008)
River Mile
Navigation Channel
River or Slough

FEATURE SOURCES:
Transportation, Property, or Boundaries: Metro RLIS.
Channel & River miles: US Army Corps of Engineers.

**DRAFT**
DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

Prepared by LSM 10/13/09 map 4000 W:\Projects\01-25-01 Portland Harbor ERA\Data\GIS\Events\Bioassay\R3\EPA_directed